1976-03

Measuring aircraft capability for military and political analysis.

LeGrow, Allan Wesley

Monterey, California. Naval Postgraduate School

MEASURING AIRCRAFT CAPABILITY FOR MILITARY AND POLITICAL ANALYSIS

Allan Wesley LeGrow

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

MEASURING AIRCRAFT CAPABILITY FOR
MILITARY AND POLITICAL ANALYSIS

by

Allan Wesley LeGrow

March 1976

Thesis Advisor: E.J. Laurance

Approved for public release; distribution unlimited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Measuring Aircraft Capability for Military and Political Analysis		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; March 1976
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Allan Wesley LeGrow		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1976
		13. NUMBER OF PAGES 161
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Scaling
Capability Measurement
Arms Transfers
Multi-Attribute Utility Scaling
Fighter Capability
Factor Analysis
Paired Comparisons Scaling
Successive Interval Scaling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis explores the question of measuring weapons capability for application in arms transfer studies and military analysis. A review of common theories and methods of scaling and a discussion of measurement techniques currently used in arms transfer research, provide background information for the sections on capability measurement.

Two conceptual approaches to capability are developed and the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(20. ABSTRACT Continued)

problems of measuring capability discussed. A discussion of possible ways to measure capability follows and four scaling techniques presented; factor analysis; paired comparisons; successive intervals; and multi-attribute utility scaling. After clarifying their theoretical bases, strengths, and weaknesses, each method is used to scale aerial combat capability in fighter aircraft.

One major conclusion reached is that judgemental scaling techniques are presently more valuable for measuring capability than more computerized procedures such as factor analysis. A second conclusion is that multi-attribute utility scaling affords the best opportunity for ratio comparisons of weapon capability.

Measuring Aircraft Capability for
Military and Political Analysis

by

Allan Wesley LeGrow
Lieutenant, United States Navy Reserve
B.A., Bowdoin College, 1970

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF ARTS IN NAVAL INTELLIGENCE

from the

NAVAL POSTGRADUATE SCHOOL
March 1976

ABSTRACT

This thesis explores the question of measuring weapons capability for application in arms transfer studies and military analysis. A review of common theories and methods of scaling and a discussion of measurement techniques currently used in arms transfer research, provide background information for the sections on capability measurement.

Two conceptual approaches to capability are developed and the problems of measuring capability discussed. A discussion of possible ways to measure capability follows and four scaling techniques presented; factor analysis; paired comparisons; successive intervals; and multi-attribute utility scaling. After clarifying their theoretical bases, strengths, and weaknesses, each method is used to scale aerial combat capability in fighter aircraft.

One major conclusion reached is that judgemental scaling techniques are presently more valuable for measuring capability than more computerized procedures such as factor analysis. A second conclusion is that multi-attribute utility scaling affords the best opportunity for ratio comparisons of weapon capability.

TABLE OF CONTENTS

INTRODUCTION -- 7

I. BASIC THEORY AND METHODS OF SCALING AND MEASUREMENT ---------------------------- 10

II. CURRENT WAYS OF MEASURING ARMS TRANSFERS --- 21

III. DEFINING CAPABILITY IN FIGHTER AIRCRAFT ---- 44

IV. POSSIBLE METHODS FOR SCALING CAPABILITY ---- 74

A. THE FACTOR ANALYTIC APPROACH ----------- 75

1. Rationale -------------------------- 75

2. Theory ----------------------------- 76

3. Scaling Aerial Combat Capability With Factor Analysis --------------- 84

B. THE METHOD OF PAIRED COMPARISONS ------- 96

1. Rationale -------------------------- 96

2. Theory ----------------------------- 98

3. Scaling Aerial Combat Capability Using Paired Comparisons ----------- 101

C. THE METHOD OF SUCCESSIVE INTERVALS ----- 108

1. Rationale -------------------------- 108

2. Theory ----------------------------- 109

3. Scaling Aerial Combat Capability Using Successive Intervals --------- 112

D. THE MULTI-ATTRIBUTE UTILITY APPROACH --- 116

1. Rationale -------------------------- 116

2. Theory ----------------------------- 119

3. Scaling Aerial Combat Capability With Multi-Attribute Utility Functions -------------------------- 129

V. CONCLUSIONS ---------------------------------- 138

APPENDIX I -- 141

APPENDIX II --------------------------------------- 144

APPENDIX III -------------------------------------- 146

APPENDIX IV --------------------------------------- 154

BIBLIOGRAPHY -------------------------------------- 157

INITIAL DISTRIBUTION LIST ------------------------- 161

INTRODUCTION

Writing in 1969, George Thayer expressed his concern over the age of arms and the total grip it has on the world.

> We live in an age of weapons. Never before in the history of mankind have the weapons of war been so dominant a concern as they have been since 1945. Armaments now have enough destructive power to destroy most life on earth. Their acquisition or presence, in a large part, determines the makeup of governments, the course of foreign policy, the thrust of economic effort, the social climate in which man lives. No significant act of contemporary history is free of their influence. Few other concerns in the world demand so much effort, time, and money.[1]

Seven years later, Thayer's comments still apply. The transfer of arms is a major activity in the international arena and continues to grow in magnitude.

Until recently, little formal effort has been given to the study of this important problem. As a result, little useful empirical theory has been developed which can be used to explain the effects of arms transfers and guide decision makers in the formulation of policy. Since 1969, however, progress has been made. What started with simple descriptions of arms flows throughout the world has now evolved into a fairly sophisticated analytic effort aimed at simplifying the complexities of the phenomenon, uncovering regularities, and developing theories.

---

[1]Thayer, George, The Arms Business. The International Trade in Armaments, p. 18, Simon and Schuster, 1969.

Despite this impressive growth, theoretical development has been sporadic because of a lack of progress in developing indicators and techniques which meaningfully measure arms flow. Money, numerical accounting of weapons by type, and military utility all have been used to operationalize arms transfers. Most of these approaches are flawed, however.

This thesis attempts to deal with the problem of operationalization by demonstrating several different measurement techniques which have potential utility in arms transfer research. Of necessity, the process begins modestly. In Chapter I, the reader is introduced to basic measurement theory and familiarized with important concepts and terminology. In addition, the importance of measurement is discussed particularly as it relates to military and political analysis.

Chapter II acknowledges the major attempts to operationalize arms transfers and tries to clarify the strengths and weaknesses inherent in each. Particular emphasis is placed on past attempts to measure the qualitative differences in arms using factor analysis. Although critically appraised, it is maintained that the factor analytic approach is valuable because it strives to develop a meaningful way to compare and evaluate military capability. A closing argument is made for the importance of capability assessment and its value to the military and political decision maker is stressed.

Following Chapter II's strong endorsement of the capability approach to operationalizing arms transfers, Chapter III explores the complexities of capability analysis using fighter aircraft as a model. Two conceptual views of capability are presented: one based on weapons performance characteristics; and the other based on a multidimensional evaluation of the weapon, operating environment, and operator skill. While no attempt is made to argue in favor of one approach over the other, care is taken to accentuate the strengths and weaknesses of each in relation to capability measurement.

Taking the two conceptual definitions developed in Chapter III, Chapter IV presents four scaling techniques and applies them to capability assessment. Because a variety of disciplines are represented, the reader is provided with the rationale for using each particular approach and the theoretical premises behind each. A major contention made is that judgemental scaling techniques merit strong consideration as a means to operationalizing capability.

Chapter V serves to review the major conclusions of the research and suggests problems that need to be addressed in future studies.

I. BASIC THEORY AND METHODS OF SCALING AND MEASUREMENT

This section will provide definitional and background information that will serve the reader throughout this thesis. It begins with a discussion of the orthodox theory of measurement. The concepts of nominal, ordinal, interval, and ratio scaling are then investigated, with particular emphasis placed on the level of information assumed in each, their interrelationships, and their limitations. Finally, the reasons why measurement is important for the analysys of military problems are explored in some detail.

The classical view of measurement, which receives its fullest expression in the works of N. R. Campbell, is more restricted than the accepted view advanced in current references. For Campbell and other classicists direct or "fundamental" measurement is possible only when the axioms of additivity are isomorphic with the manipulations performed upon objects.[2] To illustrate, measurement can be associated with length because the addition of two lengths results in a third length whose magnitude equals the sum of the first two and which makes sense within the context of the operation.

---

[2] Churchman, C.W. and Ratoosh, Philburn (ed.), Measurement Definitions and Theories, p. 22, John Wiley & Sons, 1959.

Adhering to such a constraint would mean that the only other things which could be measured (aside from length) would be time and mass. To allow for quantification of other important phenomena, classicists recognized an indirect or derived form of measurement in which magnitudes are defined through laws relating fundamental magnitudes. Density, the ratio of mass to volume (length), is thus an example of indirect measurement. Notice that the addition axiom does not hold since adding two substances with equal densities does not produce a substance with twice the density. The one aspect which does link direct and indirect measurement, though, is the implicit understanding that measurement makes sense only when the numbers have a direct physical interpretation.

The classical view of measurement was challenged in 1932 when a panel of distinguished British scientists discussed the feasibility of quantitatively estimating sensory events. Conservatives argued against the proposal because to accommodate it a new, more general theory of measurement would have to be accepted. "Why," complained Campbell, "do not psychologists accept the natural and obvious conclusion that subjective measurements of loudness in numerical terms (like those of length...) are mutually inconsistent and cannot be the basis of measurement?"[3]

---

[3] Ibid., p. 22.

While physicists and others scoffed at the notion of measuring the subjective, the growth of psychology demanded the precision only quantification could provide. In response to this demand, S.S. Stevens proposed four scales of measurement in 1946 which were differentiated by the number and type of mathematical transformations that left each scale invariant. The greater the number of transformations that could be applied to a scale without altering its structure, the less precise the scale. The four scales, nominal, ordinal, interval, and ratio, all have endured the test of time and form the basis of modern measurement theory.[4]

The nominal scale is the "lowest" of the four scales in the sense that no assumptions are made about the values being assigned to the data. Any one-to-one mathematical transformation (which includes all of those mentioned in this chapter) can be applied without distorting any information. Values serve merely as labels for distinct categories and cannot be used to rank-order data points or measure the distance between them. Nominal scale values, in other words, are symbols which indicate common class membership only, and hence, cannot be added, multiplied, or manipulated in any other way. It follows, therefore, that statistical techniques

[4]Other scales have been developed, most notably the logarithmic interval scale and the ordered metric scale. Because of their infrequent use they are not considered, however.

which depend on the distance between data points or on its order, e.g., mean, median, standard deviation, should not be used to describe relationships between nominal data.

An ordinal scale and ordinal-level measurement results when the data can be ranked according to some criterion. Since the only thing that must be preserved is the rank-order of the data, any monotone increasing transformation can be used to create different scale values if desired without distorting the scale. While empirical operations of "greater than" and "less than" can be accomplished, nothing can be said about the distance between data points or categories. Thus, ordinal scale values order the data and indicate relative magnitude along a continuum, but they do not exhibit any other properties of the real number system.

If, in addition to ordering, distances between categories can be defined in terms of fixed and equal units, interval measurement is possible. Linear transformations of the form $x' = a + bx$, $b > 0$ (where x' is the transformed scale value, x is the original scale value, and a and b are real numbers) can be used to adjust scale values when appropriate. Since distances between points can be calculated with interval values, some of the strongest statistical tools, e.g., standard deviation, product-moment correlation, and factor analysis, can be used to describe the data or to advance theory. In other words, interval measurement allows more subtle relationships to be explored than is possible with ordinal and nominal measurement.

Finally, the highest level of measurement is embodied in the ratio scale. Being the highest, its structure remains invariant only with similarity transformations of the form $x' = bx$, $b > 0$ (where x' is the transformed value, x is the original scale value, and b is a real number). The ratio scale retains all of the properties of an interval scale with the additional feature of a natural, fixed zero point. This is significant because it allows for ratio comparisons between data points. In sum the values on a ratio scale exhibit all the properties of real numbers and afford the greatest flexibility for describing and reporting relationships within a set of data.

A summary of the properties of Steven's four scales can be found in Table I.

Three consequences of Steven's scales are worth emphasizing at this point. First the range of operations possible, from basic classification at the nominal level, to the extensive mathematical manipulation possible with ratio values, requires a general definition of measurement. For the purpose of this study, therefore, measurement will be considered as the assigning of numbers to objects (data) according to a set of rules. The rules can be as general or restrictive as the circumstances and goals of the research demand, but whatever the case, they must be applied consistently.

Second, the level of measurement required depends on the research questions asked. If the analyst wants to know, for example, whether arms transfers to a certain area increase

TABLE I[5]

Scale	Characteristics	Transformations Leaving Scale Form Unchanged	Examples
Nominal	Classification or grouping. Determination of equality is possible	Any 1-1 transformation	Numbers on football players
Ordinal	Order or ranking where instances are assumed to be on a continuum. Determination of equality and of greater or less is possible	Any monotone increasing transformation	Military Rank
Interval	Arbitrary base (origin). Arbitrary unit which is constant over the scale. Determination of (1) equality, (2) greater or less, and (3) the equality of ratios of intervals or differences	Any linear transformation, $a + bx$, $b > 0$	Temperature (F or C) Quality point rating Factor Scores
Ratio	Fixed or natural origin. Arbitrary unit, constant over the scale. Determination of (1) equality, (2) greater or less, (3) equality of ratios of intervals, and (4) equality of ratios	Any similarity transformation bx, $b > 0$	Length, Cost

[5]Adapted from Stevens, S.S. Handbook of Experimental Psychology, Wiley, 1951.

or decrease conflict, ordinal measurent is adequate. If the problem is trying to ascertain the degree to which conventional arms transfers influence conflict, ratio measurement is needed. It is generally felt that the higher the level of measurement the better off the researcher will be. Assuming the time and resources are available, this is a sound operating principle.

Third, the use of a particular mathematical model or statistical technique to describe the data is governed by the level of measurement employed. The more powerful analytical tools are reserved for ratio and interval data. For most statisticians, this is an immutable law. H.M. Blalock, for instance, states that

> ...it is not legitimate to make use of a mathematical system involving the operations of addition or subtraction when this is not warranted by the method of measurement. Ideally, one should make use of a data gathering technique which permits the lowest levels of measurement, if these are all the data will yield, rather than using techniques which force a scale on the data.[6]

Surprisingly, this viewpoint is not accepted by all statisticians. John Tukey argues that the fact that data are collected on, for example, an ordinal scale, should not in itself restrict the analyst to low-level analysis. In fact, when using judgmental scales to quantify subjective

---

[6]Blalock, H.M., Social Statistics, pp. 17-20, McGraw-Hill, 1972.

data, there are methodological justifications for using more powerful analytic techniques, most notably, the Thurstonian laws of comparative and catagorial judgment. This has important consequences for arms transfer study since some of the measurement techniques explore in Chapter IV of this thesis are judgmental.

Tukey's philosophy is not a carte blanche to mate low-level measurement with high-level statistics, but rather an exhortation to glean all the information possible from the data. The tendency when dealing with "soft" variables is to be overly conservative. If the data is not interval, it is categorized as ordinal. As Tukey notes, however, the typical state of knowledge short of metric (interval/ratio) information is not rank-order but actually something more.[7] The analyst should be attuned to this.

It is important to realize that under certain circumstances the level at which data is measured can be changed to suit the purposes of the research. It is fairly obvious that ratio or interval data can be scaled at an ordinal or nominal level because the data possesses more than the required amount of information and precision. This is not generally done, though, since such a conversion would result in a loss of information. It may also introduce error since

---

[7] Abelson, R.P., and Tukey, J.W., "Efficient Conversion of Non-Metric Information into Metric Information," in The Quantitative Analysis of Social Problems, E.R. Tufte (ed.), p. 407, Addison-Wesley, 1970.

"artificial" boundaries are imposed on the data. Less obvious, and of more potential use to the military analyst, are those techniques which "up-grade" the measurement from nominal to ordinal or ordinal to interval levels. Changing nominal data to ordinal can be accomplished by using the Spearman rank correlation if a suitable reference variable can be found.[8] Ordinal to interval conversion is accomplished through class ranking and expected order statistics.[9] The exact procedures involved in scale conversion are covered elsewhere. At this point it is sufficient simply to know these maneuvers are possible.

Thus far, the emphasis has been on measurement theory and on elucidating the characteristics of the four most commonly used scales in social research. An obvious question to ask at this point is "Why should an analyst — and, in particular, the military analyst — be concerned with measurement in the first place?" It is important to suggest some reasons.

One very compelling reason for such emphasis is the nature of the world in which the military now operates. The problems are complex and the quantity of information immense. Computers are not just fashionable, therefore, but are absolutely essential in the analysis of problems. Being able

---

[8]Anderberg, Michael R., Cluster Analysis for Applications, p. 53, Academic Press, 1973.

[9]Ibid., pp. 56, 59.

to describe politico-military variables in a manner that is amenable to computer processing thus becomes a significant venture. Measurement helps accomplish this.

Gurr notes that the ultimate goal of almost all empirical research is the development of empirical theory.[10] This is a central concern of those investigating the effects of arms transfers in various parts of the world as well. Quantitative research (and hence measurement) contributes to theoretical development by increasing precision since the ability to precisely define phenomena allows the researcher to test theoretical statements more rigorously and with more assurance. Thus, in an important sense, a fundamental prerequisite for the continued development of arms transfer theory is measurement.

Another reason scaling and measurement are important is that they provide standardization. For a science to grow, other researchers must be able to replicate previous research, and to do this requires reliable measuring procedures. (This problem currently plagues arms research, especially when monetary data is used.) In a more practical sense, standardization is important in the military because of frequent changes in personnel at all levels and in all organizations. A related advantage is that by being standardized, a scale increases objectivity since the biases

---

[10] Gurr, Ted R., Politimetrics, p. 6, Prentice-Hall, 1972.

of each analyst are controlled and minimized. Even the biases of the individual who devised the scale are, by virtue of the openness and availability of the scale, subject to great scrutiny and control.

Lastly, scaling and measurement permits the conciseness of mathematics to be used in describing empirical relationships, expressing theories, and reporting results. Additionally, since mathematics is a universal language, measurement leads to greater potential communication and utility.

Measurement, and the quantification it leads to, will not guarantee success or improve analysis unless it is done correctly and effectively. With this in mind, the next chapter presents the most common measurement techniques used in the study of conventional arms transfers to Third World countries.

II. CURRENT WAYS OF MEASURING ARMS TRANSFERS

Despite the excellent work done by Amelia Leiss and Geoffrey Kemp as members of the Arms Control Project at MIT, the Stockholm International Peace Research Institute (SIPRI), and others, the empirical study of conventional arms transfers is still undeveloped. As mentioned previously, one factor which has deterred growth is a lack of progress in developing indicators and techniques which meaningfully measure arms flow. Meaningful measurement of conventional arms is an ambiguous notion that must be defined. In this study it will connote assigning numbers to conventional arms according to a set of rules such that the quantitative and qualitative dimensions of arms flows are represented. Such measurement should be done at the highest level possible to minimize restrictions in using the data and maximize the statistical tools available to the analyst.

In addition to these requirements, a good measurement procedure must have validity and generate pertinent information which is free of systematic and random error. In contrast to the multiple criteria that must be satisfied to prove indicator validity (e.g., face validity, convergent and criterion tests, etc.), the validity of a measurement technique depends essentially on the soundness of the theoretical principles which underlie it, as well as on its ability to provide accurate information at the required level of precision. All measurement theory, it should be remembered, espouses a set of rules and some/all of the properties of the real number system. Meeting the requirement of acceptable theory, therefore, usually poses no problem. What can be a problem is insuring that the scale and level of measurement is appropriate for the research questions asked. For example, questions requiring information about general trends in arms transfers can be accommodated with nominal or ordinal scales and indicators. However, when more precise information is required, or when the analyst wishes to operationalize arms transfers for a multiple regression analysis, interval-level measurement is required. In sum, no matter how sound the measurement theory, the validity of a measurement technique cannot be assessed apart from the specific research questions it serves.

The measurement technique should also be reliable; that is, repeated measurement of the same object/attribute should yield consistent results. While it is always true that the measuring instrument will induce some error, this can be minimized by insuring that the instructions and rules comprising the measurement procedure are sufficiently detailed and explicit and are observed religiously. This is especially important when using judgemental measurement techniques where discriminal differences among judges are commonplace. Again, the crux of reliability is replicability and consistency over a series of trials. Any technique which cannot provide such duplication should be discarded.

There are formal statistical techniques which can help the analyst measure the reliability of a scale. For ordinal scales (or ordinal ranking which is assumed to have an underlying interval measure), the Spearman rank correlation coefficient (Rho) or Kendall's tau can be used to determine the correlation of scaled results. For judgemental scales, split-halves reliability procedures are available. The analyst should make use of these simple statistical tests as a matter of course.

With these criteria in mind, three techniques presently used to measure arms transfers will be analyzed in this chapter: dollar-value measurement, numerical/inventory measurement, and capability measurement.

Traditionally, the most common technique used to measure conventional arms flow is what can be described as the dollar-value approach. Quite simply, analysts using this technique quantify the volume and direction of arms flow in terms of the dollar-value of the weapons systems. SIPRI and the Arms Control and Disarmament Agency (ACDA) are two organizations in particular which rely heavily on this approach.[11]

Admittedly, dollar-value measurement does provide expedient and often useful information, and can generally be accepted as a good first estimate of arms levels. There

---

[11]For other examples, see Lambelet (1971), Milstein (1970 and 1972), and Safran (1969).

is another positive feature in this approach, namely, that it provides ratio measurement[12] and thus affords the researcher the opportunity to compare systems through the ratio of their costs, and take advantage of all of the features of the real number system.

Dollar-value measurement of arms can also be used effectively as an indicator of commitment or alignment between nations, or as a variable in a multiple regression model. J.S. Odell's work correlating U.S. military assistance (measured in U.S. dollars) with recipient nations' economic value to the U.S., is an example of such usage.[13]

However, dollar-value measurement can be misleading for several reasons. In the first place, fluctuations in arms expenditures do not always reflect the actual magnitude of the arms transferred. Occasionally, increasing expenditures correspond to decreasing numbers of weapons actually transferred. The chief reasons for this are inflation and fluctuating exchange rates. In some reports this disparity between expenditures and the number of systems transferred is partially alleviated by adjusting yearly figures on the basis of some arbitrarily selected exchange rate. Unfortunately,

---

[12]This, of course, ignores arguments in value theory which suggest that money has different psychological value depending on the amount.

[13]Odell, J.S. "Correllate of U.S. Military Assistance and Military Intervention" in Testing Theories of Economic Imperialism, ed. by S.J. Rosen and J.R. Kurth, pp. 143-161, D.C. Heath, 1974.

fluctuations usually occur too rapidly to be compensated for by this procedure and information about a weapon's true value at the time of the transaction is obscured. This is especially troublesome when using dollar-value figures for side-by-side comparisons of several countries' arms trade.[14]

A second weakness in the dollar-value approach rests in the uncertainty with which foreign — most notably Communist — weapons are priced. ACDA admits that their valuation of Communist arms exports reflects Soviet foreign trade prices which tend to underestimate the value of the equipment in terms of Western production costs.[15] Moreover, there is insufficient information available from which to perform any systematic price adjustments. As Sivard notes, "Although statistical work on such parity rates is underway, under international sponsorship, the availability of purchasing power parities for a large selection of countries is some distance in the future."[16] Hence, there is acknowledged doubt regarding cost data for foreign arms transfers.

Third, there is no standard formula that can be used to determine the value of military equipment transferred

---

[14] World Military Expenditures and Arms Trade 1963-1973, Washington: Arms Control and Disarmament A-ency, p. 1, 1973.

[15] The International Transfer of Conventional Arms, Washington: Arms Control and Disarmament Agency, p. 1, 1973.

[16] Sivard, Ruth L., World Military and Social Expenditures 1974, p. 30, WMSE Publications, 1974.

from excess stocks or equipment that is no longer of use to a supplier country. In a report to Congress, ACDA valued all transfers from excess stocks at "approximately one-third of acquisition cost."[17] Michael Mihalka, however, cites instances where excess military equipment was transferred to other nations at less than one-tenth to one-hundredth the initial cost.[18] Clearly, in such instances it is virtually impossible to determine the real cost to the buyer and value to the seller.

Fourth, in transactions with most Third World nations it is extremely difficult to determine exactly how arms deals are financed. Leiss, for instance, notes that barter is sometimes a part of arms trade.[19] Political, economic, and other concessions may also be involved which, although not reflected in published prices, would have a direct but indeterminable relationship to the value of the weapons (e.g. the Soviet naval base at Mursa Matruh in Egypt). Additionally, internal factors common to many Third World countries such as corruption, inefficiency, and incompetence would distort the figures even more. The analyst clearly

---

[17] Ibid., p. 20.

[18] Mihalka, Michael, Understanding Arms Accumulation: The Middle East as an Example, p. 14, University of Michigan (Mimeo), 1973.

[19] Leiss, Amelia, et. al., Arms Transfers to Less Developed Countries, C/70-1, p. 31, MIT, 1970.

must be cautious when drawing inferences from foreign arms expenditure data.

Finally, money is an unreliable indicator of the qualitative differences in arms. While it is generally true that the cost of arms is directly related to their level of sophistication, and in this sense to their quality,[20] the factors discussed previously distort this relationship. Even if it could be assumed that the more costly the weapon the more sophisticated it is, this does not necessarily relate to the military value of the system. Mihalka [1973] points to the F-111 and the C-5A as examples of very costly weapons systems whose military value has been relatively low. In terms of arms trade to lesser developed countries, this problem is further exacerbated by uncertainties in the capacities of the countries to absorb, maintain, and effectively use sophisticated weapons and by a general lack of knowledge about spare parts and required training.

To summarize, the problem with the dollar-value approach is not the level of measurement, but the instability of the measurement unit. Reliability is a serious problem. In many instances cost does not even depict the volume of arms transferred accurately. What is more, it fails to measure qualitative differences in weaponry and, as a result, cannot be used to measure military balance or potential. Cost can

---

[20]A good example of this is the Escort Ship Cost Model (ESCOMO) developed by R. Wilson at the Center for Naval Analyses.

be used as a variable in a regression model or as an indicator to explain other phenomena, but in view of the serious reliability problems, and the lack of standardization in its usage, it would only have limited utility.

Most analysts relying on dollar-value measurement are well aware of its fallacies and usually catalog their reservations with their analyses. Some researchers, such as Amelia Leiss, have taken a more positive approach to the problem by using an alternative unit of measurement, the weapon system itself. In terms of usage, this type of measurement — referred to here as the numerical/inventory approach — is almost as popular as the dollar-value approach.[21] Insofar as the numerical inventory method reflects actual weapons amounts transferred, the technique unquestionably improves measurement reliability. For example, it is able to provide a more accurate accounting of the volume and direction of arms traffic, since, by focusing on the actual weapons themselves, many of the factors which distort dollar measurement (such as arbitrary exchange rates and inflation) are eliminated.

Numerical/inventory measurement cannot be used effectively to describe qualitative differences in arms transfers because the unit of measure (the weapon itself) does not embody any

---

[21]See, for example, Leiss [1970], Kemp [1970], and SIPRI [1968-1974].

general attribute or characteristic upon which to base comparisons with other weapons. The most successful measurement of qualitative difference in arms in Leiss' work, for example, occurs only when she constructs indices based on some inherent weapons characteristic. The best example of this is her "modernity index" which measures the "modernity" of aircraft on an interval scale.[22] It appears that if the qualitative aspects of arms must be measured, some abstract attribute or characteristic must be identified and agreed upon as the basis for comparison.

The fact that the numerical/inventory approach cannot in itself represent qualitative differences in arms makes it inappropriate as a measure of military balance or capability — an issue of vital importance to military analysts and decisionmakers. As the 1967 Arab/Israeli war and the Vietnam rout in 1975 both show, numerical superiority in armament does not equate directly to military strength. This is perhaps an obvious point, and yet it is frequently forgotten or distorted in arms studies because most comparisons of the military strengths of countries are in fact done on the basis of numbers of systems. The inevitable implication is that the more weaponry a country has, the better off it is. (Consider The Military Balance Series published by IISS which purports to be "a quantitative evaluation of the

[22]Leiss, Amelia, Changing Patterns of Arms Transfers: Implications for Arms Transfer Policies, C/70-2, pp. 19-35, 219, MIT, 1970.

military power ... throughout the world" and is based solely on a tabulation of men and equipment.) A definite need exists for measurement schemes which can cope with the qualitative factors of arms analysis.

A promising and sophisticated attempt to measure the quality of conventional arms is presented by Michael Mihalka. In search of a better indicator than military expenditures and numerical inventories to measure arms accumulation, Mihalka proposed to measure weapon system capability. The main assumption behind such measurement is that any weapon can be viewed as a linear combination of its component capabilities, each of which can be measured at an interval level. Thus, by selecting the appropriate performance characteristics, it is possible to derive a numerical value which reflects total system capability. A related assumption is that all the characteristics of a system will reduce to two underlying dimensions or factors, offensive and defensive. Intuitively, this is quite pleasing since military planning and tactics are dichotomized the same way. In view of the salience of capability as an attribute of arms, and the prospects of interval measurement, it is profitable to examine Mihalka's methodology in some detail.

To begin with, selecting appropriate performance variables is based on three considerations: (1) the number and type of variables deemed necessary to define the weapon system adequately; (2) whether or not the variable is amenable to quantification; and (3) the availability of data.

The second consideration is particularly striking because it implies that capability is an inherent attribute of a weapon and is satisfactorily described by "hard" quantifiable, performance data. (Commitment to this viewpoint has a profound effect on the acceptability of several of the measurement techniques proposed later in this study.)

The analytical model used by Mihalka to measure the variables is the oblique multi-dimensional factor model which identifies main factors by grouping mathematically related input variables into distinct clusters.[23] Mihalka's analysis of selected variables for aircraft results, for example, results in two groupings: (1) speed, technological data, and performance — defining the defensive factor; and (2) payload and combat radius — comprising the offensive factor. The following data, taken from Mihalka's study of arms accumulation in the Middle East, describes these clusterings in terms of raw factor loadings. Since the loadings are the correlation coefficients between variables and factors, the higher the value, the greater the correlation between factor and variable.

---

[23]Rummel, R.J., Applied Factor Analysis, p. 409, Northwestern University, 1970.

TABLE 1[24]

Variable	Factor I (Defensive)	Factor II (Offensive)
Speed	0.917	0.015
Tech date	0.723	0.026
Performance	0.848	- 0.051
Payload	0.161	0.918
Radius	- 0.258	0.719

Since each factor defines a group of interrelated characteristics, it can be considered a functional unity and used as a scale.[25] Factor scores are derived in the following way. Each variable is weighted according to the statistical variation it has in common with the offensive and defensive factors and multiplied by the data value of each case. The sum of these weight-times-data products for the three variables comprising the defensive dimension yields the defensive factor score. Similarly, the sum of the weight-times-data products for payload and radius leads to the offensive factor score. A sample of the aircraft scores obtained by Mihalka is found in Table II.

---

[24]Mihalka, op. cit., p. 21.

25Rummel, op. cit., p. 30.

TABLE II[26]

DERIVED CAPABILITY SCORES
FOR AIRCRAFT

Name of System	Manufacturer	Capability Scores Defensive	Capability Scores Offensive
1. Interceptors			
F-104A Starfighter	US	1.410	- 0.280
Lightning	UK	2.554	- 0.767
MiG-21 D/F	USSR	1.645	- 0.795
2. Multipurpose Fighters			
F-4 C/D	US	1.695	1.106
F-5A	US	0.885	0.040
MiG-19	USSR	0.997	0.571
Mirage V	France	1.561	0.311
Mystere IV A	France	0.873	- 0.593
3. Strike Fighters			
A-4F	US	0.631	0.787
SU-7B	USSR	1.342	- 0.318
4. Light Bombers			
IL-28	USSR	- 0.439	2.016
5. Medium Bombers			
Tu-16 A/B	USSR	0.473	3.737

Notice that a definite pattern exists in the scores. Interceptors are characterized by high speeds and high performance and thus have high defensive scores. Logically, the system with the highest score has the greatest capability — in this case, the British Lightning. Bombers score high on the

[26]Mihalka, op. cit., p. 18.

offensive dimension because of their typically high payload and range. Strike aircraft and multipurpose fighters generally score in the middle ranges of both dimensions.

After deriving capability scores for individual systems, Mihalka takes the analysis one step further and combines them with information on the numerical inventories of a subset of Middle East countries to obtain capability indices for each country's weapons stockpile. The precise relationship is given by

$$C_{i,k,t} = F_{i,k,s} * I_{s,t}$$

where _s_ denotes a particular weapons system, _t_ the time, _i_ the class of system (e.g. air, ground, naval), _k_ the factor, _F_ the factor score, _I_ the numerical inventory, and _C_ the capability score of the inventory. Mihalka's results are provided in Table III for illustrative purposes (see page 35).

At first glance, using factor analysis to measure capability seems to be a promising technique for arms transfer analysis. For example, when Mihalka uses capability scores as variables in a linear model to predict the military capability of several Middle East countries, the explained variance, R^2, ranged from 0.65 to 0.93.[27] This indicates a

---

[27]_Ibid._, pp. 31-42.

TABLE III[28]

DERIVED AIRCRAFT CAPABILITY INVENTORIES FOR MIDDLE EASTERN SUBSET

Year	Iraq	Israel	Jordan	Syria	Egypt
49	0.0	0.0	0.0	0.0	34.87
50	1.640	0.02	0.0	0.0	49.74
51	3.110	0.04	0.0	0.0	63.18
52	4.440	0.06	0.0	0.0	75.17
53	5.640	25.64	0.0	4.380	67.65
54	32.86	23.09	0.0	8.320	60.88
55	33.95	151.2	0.0	37.42	507.8
56	37.87	305.9	15.76	207.5	451.8
57	92.44	275.3	14.18	191.1	584.8
58	131.5	247.8	40.05	172.0	526.3
59	136.3	223.1	36.05	154.8	473.7
60	242.1	343.8	36.08	150.1	453.0
61	282.3	309.4	32.46	145.8	726.0
62	265.3	278.5	29.22	142.0	1100.0
63	244.9	462.7	26.31	138.6	1017
64	304.8	416.4	23.67	135.5	942.4
65	298.4	394.3	21.29	121.9	1013
66	401.6	382.8	19.18	258.9	970.0
67	501.2	424.9	25.62	317.5	1105
68	617.7	468.5	136.8	285.7	1049

[28] *Ibid.*, p. 22.

certain degree of predictive validity since the closer R^2 is to 1.0, the more successful the model is at predicting the dependent variable. Another positive feature is the interval measurement of capability, not only because of the achieved measurement level but because of the salience of capability as a unit of measure as well. Successfully measuring capability would provide the common denominator needed to compare weapon systems and address the critical question of military balance.

However, there are some significant problems with Mihalka's methodology. Consider, first, the process leading to the aircraft inventory capability scores (Table III). The first step of this calculation involves adjusting the derived weapons capability scores (Table II) so that there are no negative or zero values. Mihalka accomplishes this by adding 0.1 (selected arbitrarily) to the absolute value of the lowest factor score and adding the resulting sum to each aircraft score. The effect is to move each system in a positive direction along the interval scale by the same amount. Recall that this is permissible with interval measurement since the information is preserved by a linear transformation. Multiplying these adjusted values by varying inventories to obtain composite country scores is tenuous, however, because the interval nature of the data is violated. An example will illustrate this. Suppose the derived factor score for aircraft A is 2.0 and for aircraft B, 1.0, along the offensive dimension. If a country had an inventory of

25 A's, the country capability score would be 25 * 2.0 = 50.0. Similarly, if a second country had 50 B's, it's capability score would also be 50.0. Clearly, this would be a situation of parity. Now consider the transformation of the individual factor scores by an arbitrary value of 0.5, i.e., aircraft A = 2.5 and aircraft B = 1.5. Multiplying these adjusted capability scores by the same country inventories yields a capability score of 67.5 for the first (25 * 2.5 = 67.5), and 75.0 for the second (50 * 1.5 = 75). A situation of equality has suddenly become an advantage for the second country without any change in the number or type of weapons. Since Mihalka used the country capability scores to determine his residuals, his results are tenuous and may not be reliable. Unless Mihalka, or for that matter any analyst using this approach, can justify a <u>varying</u> multiplicative transformation on <u>interval</u> data, this will always be the case.

An additional problem is the necessary assumption that the selected variables adequately describe the attribute being measured along each of the dimensions. Many analysts do not pay enough attention to variable selection, or they justify including particular variables solely on the basis of high factor loadings. Unfortunately, even a very high loading (correlation) between factor and variable does not assure proper variable selection. The high nonsense correlation presented by British statisticians Yule and Kendall between the growing number of radios in the U.K. and the growing number of mental defectives with correlation, $R = .99+$, and explained variance, $R^2 = .99$, is vivid proof of this.[29] Thus, unless one is comfortable with the assertion that speed, technological date, and performance (which, incidentally, is not defined by Mihalka) equate to "defensive capability", the results are debatable.

A related and equally important concern is the validity of the factors themselves. As Gurr points out:

> Factor analysis results always pose problems of interpretation. A fundamental source of dispute is whether the factors are merely useful artifacts of the analysis or whether they represent latent but real phenomena.[30]

The realist position predominates at the present time and is accepted by this author. Oftentimes the reluctance of some people to accept the validity of factors can be traced to poor factor-labelling more than the actual clustering of variables to create dimensions.

A second attempt to measure arms in terms of capability was advanced by Lewis Snider in 1975.[31] Aside from a difference in labelling of the two factors (air-to-air combat

---

[29] Tufte, Edward R., Data Analysis for Politics and Policy, p. 88, Prentice Hall, 1974.

[30] Gurr, *op. cit.*, p. 157.

[31] Snider, Lewis, Middle East Maelstrom. The Impact of Global and Regional Influences on the Arab-Israeli Conflict. 1947-1973, Ph.D. Thesis, University of Michigan, 1975.

and ground attack vs. defensive and offensive), the methodology is the same as Mihalka's. Snider, however, derives twelve key variables instead of five as Mihalka did.[32] The clustering of these variables and their loadings are reproduced in Table IV.

TABLE IV[33]

FACTOR ANALYSIS OF PERFORMANCE CHARACTERISTICS OF COMBAT AIRCRAFT

Variable Name	Factor I (Air-to-Air Capability)	Factor II (Ground Attack)
Year production began	.97	.12
Speed: primary mission	.99	.02
Maximum speed	1.00	.00
Service ceiling	1.00	-.11
Horsepower or Thrust	.97	.11
Rate of climb	1.00	-.14
Maximum take-off weight	.16	.95
Ordnance payload	.16	.95
Ferry range	-.13	1.00
Combat range	.00	1.00
Combat radius: external fuel	.03	.99
Combat radius: internal fuel	-.20	1.00

[32]It is assumed that Snider and Mihalka both started with more than 12 and 5 variables respectively although the precise numbers are unknown by the author.

[33]*Ibid.*, p. 251.

This is a more comprehensive selection than that offered by Mihalka, but it is still open to criticism. For instance, thrust-to-weight ratio (T/W), wing loading (W/S); specific excess power (P_s), turn rate, and turn radius are the factors which are used to describe maneuverability and aerial combat capability in authoritative sources such as AGARD Conference Reports, the Navy Tactical Manual for F-4 Aircraft, the Journal of Space/Aeronautics, and Aviation Week and Space Technology. Expressed in other words, aerial combat capability depends on both the potential energy of the system and its maneuverability. It is clear that Snider's variables relate only to energy factors with no regard to maneuverability factors. Hence, the systems which score highest are brute-force aircraft such as the MIG-25 FOXBAT, while less powerful, more maneuverable systems such as the F-16 have a lower scale value. However, most military commentators would not accept the conclusion that the F-16 is inferior to the MIG-25 in air-to-air combat capability. Again, factor analysis demands that the variables define what the analyst intends to measure along the derived dimensions as precisely and completely as possible. If the analyst cannot be sure of this, he may be measuring something other than what is intended.

Although Snider does not calculate country capability scores, he does calculate a composite capability score which he calls the "arms transfer score". This is obtained by multiplying the product of the two factor scores for each particular weapon by the number of weapons transferred. Once again it was first necessary to adjust the individual capability scores to eliminate zero and negative values. In contrast to Mihalka's conversion formula, Snider adds a value large enough to transform the highest negative score to 1.00.[34] For reasons previously discussed, both manipulations have a profound effect on the scale values and hence on the conclusions and results of Snider's analysis.

Hopefully, the inadequacies of the three measurement techniques presented in this chapter have been clarified. Different weaknesses plague each one. Dollar-value measurement is perhaps the least reliable measure of arms flow and military potential, and yet, ironically, is the approach most often used in arms research. Often it cannot provide reliable information for even the most fundamental questions — such as determining the magnitude of arms flow — due to fluctuations in the value of the dollar. Additionally, it is susceptible to differential exchange rates, over/under-valued exchanges due to ancillary conditions, and is further constrained by its inability to reflect the impact of arms in relative terms. Numerical/inventory measurement, on the other hand, provides fairly reliable ratio measurement of arms volumes and trends and has been used successfully for descriptive analysis. Yet it is misleading when used to

---

[34] Ibid., p. 254.

represent military balance or capability. Moreover, it cannot be used to compare the relative strengths of weapon systems. Finally, while capability measurement using factor analysis does provide a basis for comparing weapons and assessing regional military balance, it breaks down in certain applications, most notably when composite capability scores for a country or groups of weapons are derived from individual weapon scores. There are also uncertainties that must be faced in selecting (or omitting) crucial variables and in interpreting the factor-dimensions themselves.

As indicated previously, each of the three measurement techniques considered has utility in certain circumstances. However, for studies of nations and regions where military factors and questions of military balance are of overwhelming importance, such as the Middle East, capability measurement is potentially the most promising of the three. It is important for other reasons as well. For one thing, intelligence analysts are always faced with the problem of assessing an antagonist's strengths and weaknesses. Capability measurement, if done meaningfully, could provide valuable inputs to improve such estimates. Secondly, such an evaluative technique could be used to evaluate U.S. weaponry, not only by facilitating side-by-side comparisons with Soviet equipment, but also through refining the selection and evaluation of new systems. Finally, it could assist the policymaker by providing an authoritative input into the arms transfer decision process. Knowing the relative strength of U.S. systems would provide some guidance as to the size and type of transfer required in a particular situation.

In view of the potential value of capability measurement to both arms transfer studies and military intelligence estimates, the remainder of the thesis will concentrate on several possible approaches to the problem. Successful measurement of capability depends, in the most fundamental sense, on operationalizing (defining) capability in a meaningful way. Is it an attribute defined only by the performance characteristics of the weapon, or is it area-dependent and affected by external factors? The next chapter will explore this issue in more detail.

III. DEFINING CAPABILITY IN FIGHTER AIRCRAFT

The analyst can view weapons capability in one of two possible ways: as an intrinsic attribute of the particular weapon expressed through the performance characteristics of that weapon; or as a more complex phenomenon composed of both intrinsic performance characteristics and external factors like the operating environment, operator proficiency, and the technological capacity of the user. The first approach can be seen in the factor analytic method employed by Snider and Mihalka where only performance variables are combined to obtain system capability scores. Among other things, the most significant assumption behind such an approach is that system capability is invariant and that additional forms which might effect a weapon's capability in a given situation need not be considered. The second approach has not been applied to capability analysis *per se* in any study familiar to this author. Geoffrey Kemp's work on classifying weapons systems and force designs comes closest to incorporating this multi-faceted view inasmuch as he considers environmental and regional factors to have a direct impact on weapons effectiveness.[35] Obviously, allegiance to this view means that capability measurement would be area

[35] See especially, Kemp, Geoffrey, *Classification of Weapons Systems and Force Designs in Less Developed Country Environments*, C/70-3, MIT, 1970.

dependent, with a system's relative potential fluctuating according to outside influences.

With these comments in mind, the main objective of this chapter is to derive definitions for the aerial combat capability of fighter aircraft using both approaches. An analysis of pertinent aerodynamic equations, along with weapons effectiveness studies by Kemp and others is performed to identify the required variables. In the process, the complexity of capability analysis is demonstrated and solutions to some data acquisition problems provided. No attempt is made to argue in favor of either of the philosophical views. Rather, it is hoped that the strengths and weaknesses of each are brought to light, particularly as they relate to scaling and measurement.

Recall that both Snider and Mihalka used the same three criteria to determine the relevance of variables. With minor modifications, they are: (1) the ability to describe the system meaningfully; (2) ease of quantification; and (3) the availability of data. All are sound, realistic principles and serve the variable selection process quite nicely. The most constraining of the three is, of course, data availability. It is an unfortunate (although sometimes necessary) fact that much of the important information needed for capability assessment is classified and thus unavailable to the research community at large. Even more unfortunate is the compromising effect such information gaps have on what should be the most important requirement — a meaningful representation of the system. For aircraft this means, at a minimum, considering the aerodynamic qualities of the airframe, the weapons suit typically employed, and its endurance (range).

Authoritative sources such as the Navy's Tactical Manual for the F-4, and AGARD (Advisory Group for Aerospace Research and Development) Conference Reports, describe the aerodynamic requirements of fighter aircraft in terms of two qualities — energy and maneuverability.[36] A plausible way to uncover meaningful variables, therefore, is to consider the aerodynamic equations describing these phenomena. First, with respect to energy, the total energy of a system is the sum of its potential and kinetic energy. For an aircraft in flight this is represented by:

$$E = mgh + \frac{1}{2} mv^2 \qquad (1)$$

where m is the mass of the aircraft, g is the force of gravity, h is the altitude, and v is the aircraft's velocity. Frequently, system energy is expressed as specific energy, E_s, which is simply Equation (1) divided by the system weight, W. That is,

$$E_s = \frac{E}{W} = h + \frac{v^2}{2g} \qquad (2)$$

---

[36]See especially, AGARD Conference Proceedings No. 62, Preliminary Design Aspects of Military Aircraft, Harford House, March 1970, and McDonnell Douglas' Tactical Manual Navy Model F-4B and F-4J Aircraft (U), NAVAIR 01-245 FDB-1T, pp. 1-2 - 1-3, 1 March 1971.

Whatever expression is used, it is clear that the key variables are altitude, h, and velocity, v. Increasing altitude and holding velocity constant, for instance, would produce an increase in total system energy. A similar increase in system energy results when velocity rises. Understanding this simple relationship is important since aerial combat is largely concerned with maintaining as high an energy level as possible. Given this maxim, the importance of speed and effective combat ceiling relative to capability is legitimized.

In addition to speed, the air superiority system must possess excellent acceleration capability within its flight envelope. In level flight, the available logitudinal acceleration, A_x, is given by,

$$A_x = \frac{g(T - D)}{W} \tag{3}$$

where g once again is gravitational force, T is system thrust, D is system drag (or resistance), and W system weight. Note that the ability to accelerate longitudinally does not involve velocity at all, depending instead on thrust, drag, and weight. This suggests that relying on velocity alone to represent fighter capability is inadequate. It should be clear that the greater the ratio $\frac{T - D}{W}$, the greater the system's ability to accelerate. Much of the U.S. emphasis in smaller, light-weight, aerodynamically clean, fighter designs (e.g., the F-16, F-17, and F-18) reflects a desire to improve present capability in this area, and substantiates including the thrust - drag - weight relationship in the variable list describing fighter capability.

Both velocity and the $\frac{T - D}{W}$ ratio are brought together in the expression for specific excess power, P_s, a quantity which can be thought of as the system's ability to change energy levels. (It also serves as the expression for the rate of climb.) More precisely,

$$P_s = \frac{T - D}{W} V \qquad (4)$$

(with all symbols defined previously). This is the ideal expression to use for comparing aircraft energy capabilities since it accounts simultaneously for velocity and acceleration. (U.S. fighter pilots, in fact, use P_s data to determine areas of definite energy advantage and disadvantage for their own aircraft in relation to major opposition platforms.) Unfortunately, P_s data for most combat systems are difficult to find in open sources, and in some cases even remain incomplete in classified literature. The main problem, in most cases, is accurately determining the magnitude of drag, D, at various altitudes, airspeeds and configurations. Two solutions are possible: (1) eliminate P_s comparisons and use thrust-to-weight ratio (T/W) and velocity as two separate variables; or (2) estimate the aircraft's drag using size, shape, and wing platform information. The first solution is expeditious and, for the most part, very adequate. However, the fact that estimates of aircraft drag can be made is significant because it suggests that P_s determinations are possible, and further that data availability problems for air combat systems are less severe than they initially appear to be.

No less significant than the energy state of a fighter aircraft is its maneuverability. Maneuverability can be viewed as the ability to achieve a high rate of turn and a small radius of turn throughout the flight envelope. Rate of turn, θ_T, is defined as:

$$\theta_T = \frac{A_z^2 - g^2}{V} \tag{5}$$

where $\underline{A_z}$ is the acceleration normal to the flight path, $\underline{g}$ is gravitational force, and $\underline{V}$ is velocity. As the equation shows, for a given combat speed, the rate of turn, θ_T, depends entirely on the acceleration normal to the flight path, A_z. Redefining A_z and substituting the new expression into the turn rate equation (5) reveals several additional variables:

$$\frac{A_z}{g} = q \frac{C_L}{W/S} = \frac{L}{W} = n \tag{6}$$

thus:

$$\theta_T = \frac{g}{V} \sqrt{\left(\frac{L}{W}\right)^2 - 1} \tag{7}$$

such that q is dynamic pressure caused by the velocity of the aircraft through a certain air density, C_L , is the lift coefficient, W/S is the wing loading, the ratio of aircraft weight to wing area, L is the lift generated by the system, W is the system weight, and n the load factor. In other words, to produce high normal acceleration, A_z, and a high turn rate, the system must be able to generate lift efficiently in relation to its weight. Knowing the particular load factor values (n) for two aircraft under comparative conditions would provide this information. However, these values are not normally available in unclassified sources because lift data can only be derived from a technical analysis of the aircraft and as such usually goes beyond the scope and purpose of open source material. However, as equation (6) reveals, a reasonable indication of this capacity is given by the wing loading, W/S, since the lower the wing loading at a given velocity, the higher the lift-to-weight ratio (L/W) and, hence, maneuverability. Similarly, if W/S is relatively high, L/W will be relatively low as will the maneuverability of the system. This relationship justifies including the W/S ratio on the variable list defining capability as a surrogate for the turning rate, θ_T.

If the turning rate is known, a simple conversion procedure gives the radius of turn, R.

$$R = \frac{V^2}{g\sqrt{L/W - 1}} \qquad (8)$$

Since no new variables are introduced, it is actually redundant to be concerned with the turn radius when the turn rate is known. Hence, it is sufficient to use just one of these expressions to help define maneuverability. On the strength of studies done by the Northrop Aircraft Corporation which suggest that turning rate is a more critical factor in air combat than turning radius,[37] turning rate is preferred by the author.

As is the case with specific excess power data, obtaining pre-calculated turn rates and radii for a large sample of aircraft requires access to classified sources. However, reasonable estimates for the turn rate and turn radius can be made, again from dimensional information on the aircraft. (The interested reader is referred to A.W. Babister's Aircraft Stability and Control for examples of such calculations.) Again, it is important to note that information about key variables can be obtained using open sources if the researcher is willing to exploit technological information and theory.

To review quickly, analyzing the aerodynamic expressions for energy and maneuverability identified some eight important variables which are directly related to the aerodynamic capability of a fighter aircraft: velocity (V), acceleration as indicated by the thrust-to-weight ratio (T/W), specific

---

[37] Aviation Week and Space Technology, 15 April 1974, p. 47.

excess power (P_s), wing loading (W/S), turn rate (θ_T), turn radius (R), the load factor (L/W = n), and the coefficient of lift (C_L). Because of the existing interrelationships between them, however, the list can be distilled down to four items: velocity; thrust-to-weight ratio; wing loading; and turn rate. In reality, these four variables reflect the author's preference and by no means constitutes the only plausible selection. What is important is not the variables themselves but the fact that both energy and maneuverability qualities are represented. Any capability analysis of air superiority aircraft which does not accomplish this is incomplete.

Along with the aerodynamics of the platform, of major importance to aerial combat capability is the system's endurance. The swiftest, most maneuverable fighter in the world would be of little value without the necessary endurance to perform the mission. Some Soviet fighters, and the MiG-21 in particular, suffer in this respect. Measuring this characteristic poses no problem since any one of a number of range figures routinely appears in open sources including ferry range, combat range, combat radius, and specific fuel consumption (a measure of the amount of fuel burned per hour). The one preferred here is combat radius since it is calculated on the basis of the aircraft's primary mission and thus captures the essence of realistic employment of the aircraft.

The third and final aspect to consider is the basic capability of the weapons suit normally employed with the aircraft. For air combat systems this means either light automatic cannons, air-to-air missiles, or both. Because of the great similarity between aircraft gun systems throughout the world, it is doubtful that a side-by-side comparison of such systems would reveal much that would influence or alter the relative capabilities of two gun-equipped fighters. For the most part, the guns are 20-23 mm in size, having firing rates of about 1000 rounds per minute, and effective ranges of less than one mile. The impact of minor variations in these specifications would be difficult, if not impossible, to express in the capability measure of the entire fighter weapon system (i.e., the combined capability of the airframe's aerodynamics, the weapons suit, and mission range capacity). On the other hand, the presence or absence of a gun system should be considered in the assessment since the gun-equipped fighter has more destructive potential in an aerial combat situation than its gunless counterpart (assuming, of course, that the gun's added weight does not hinder maneuverability, etc.)

A bit more attention must be given to individual differences in missile systems since they are more pronounced and could make a difference in the fighter's assessed capability. Perhaps the best single indicator of a missile system's capability is its launch envelope which tells something about missile range, maneuverability, and required launch parameters. Again, finding this information in open sources represents a formidable problem. Because of this, other, more general, factors must be considered to piece together the missile's true capability.

One such factor is the guidance system. Differences in guidance schemes account for differences in firing envelopes and tactics. A beam-rider such as the Soviet-built Alkali or a semi-active homing system such as the AIM-7 built by Raytheon, for instance, can be fired at a closing target, or one that is crossing the flight path. This means that the pilot does not have to achieve a position behind the target aircraft to fire and, further, that the aircraft's maneuverability is not necessarily critical to a successful attack. In contrast, missiles relying on electro-optical guidance, e.g., heat-seekers such as the Atoll and Sidewinder, must be fired at the rear quadrant/exhaust area of the opposing aircraft and thus are totally dependent upon the aircraft's capability to maneuver into an advantageous position. This suggests that mating a particular missile with an aircraft can either place special demands on the aircraft or compensate for its deficiencies. Theoretically, for example, an A-6 attack aircraft, or even an S-3, could be used in an air-superiority role if equipped with a long-range missile like the Phoenix, because the missile does all the work and demands nothing of the aircraft except target acquisition, and transportation to the launch position.

The previous example suggests a second general characteristic that can be used to gauge missile capability – range. Having a missile that can reach a target more than 50 miles away certainly is more formidable than one with a 2-3 mile range, and the difference should be reflected in the system's assessed capability.

Finally, inasmuch as missiles depend on the same aerodynamic qualities as aircraft, knowing something about their maximum speeds and maneuverability would contribute to the task of determining how much capability they added to individual aircraft systems, especially in an aerial combat environment. A figher equipped with a highly maneuverable missile similar to the one depicted in Figure 1 (p. 56) would have a decided edge over any missile system currently in operation and enhance the probability of destroying an opponent because of superior aerodynamic qualities.

Despite the painstaking process used to isolate salient features of the air superiority system in terms of its aerodynamic qualities, range, and weapons suit, technological advances and/or design innovations which affect performance and capability cannot be captured with the variables selected. To illustrate, one of the newest U.S. fighters, the F-16, has fuselage strakes (Figure 2, p. 57) which produce vortices that improve lift and make the <u>effective</u> wing area

FIGURE 1

ENHANCED CAPABILITY WITH HIGHLY MANEUVERABLE MISSILE

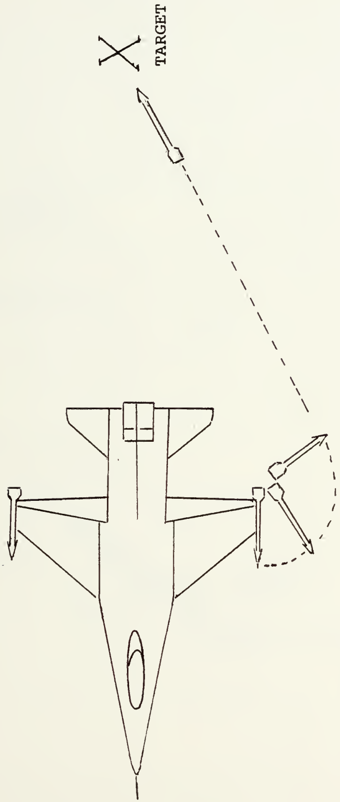
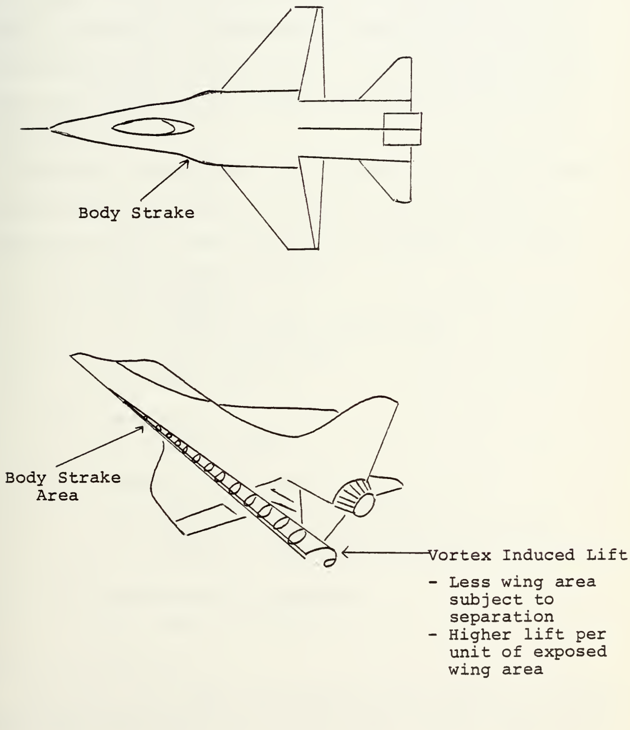


FIGURE 2

ADVANCED TECHNOLOGY OF F-16



significantly greater than the actual geometric area.[38] The calculated wing loading, W/S, is thus <u>higher</u> than the actual wing loading. It follows that using the calculated wing loading to indicate its maneuverability would lead to an underestimation of the F-16's maneuvering capability. To account for such occurrences it is recommended that the year of production, or some other time-related reference, be included in the variable list. Both Snider and Mihalka found this to be useful in their assessments of aircraft capability.

To review, using open source information and aerodynamic formulae, the following variables have been selected to define the aerial combat capability of fighter aircraft:

I. Energy/Maneuverability
   1. Velocity
   2. Thrust-to-weight ratio (T/W)
   3. Wing loading (W/S)
   4. Turn rate, θ_T

II. Endurance
   5. Combat radius

III. Weapons Configuration
   6. Guns — presence or absence
   7. Missiles — operationalized by guidance, velocity, range and maneuverability parameters

IV. Technological Innovation
   8. Year of Production

---

[38] A brief but useful discussion on the effect of strakes on fighter maneuverability can be found in AGARD Conference Proceeding No. 102, pp. 24-5, and 24-6, Harford House, 1972.

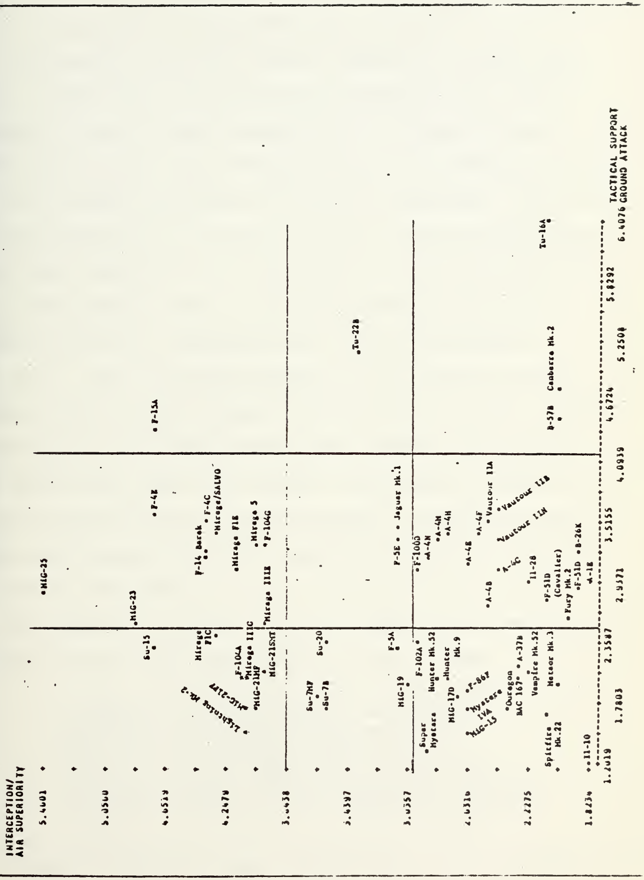
For many in the intelligence community, the technical analysis presented in this chapter is familiar. Pilots and aeronautical engineers would also feel comfortable thinking in terms of the variables presented. For others, particularly those interested in using capability to determine the political/military consequences of arms transfers, such complexity may seem unnecessary. The author is convinced, however, that without a sufficient amount of probing and technical analysis, erroneous or misleading capability assessment is highly probable. Recall the variable list Snider used to compute aircraft capability scores (Table IV, p. 39). They provide an excellent description of the energy dimension of the aircraft considered, but completely ignore maneuverability qualities. The result is to accentuate the capability of powerful, fast aircraft (i.e. pure interceptors) vis-a-vis less powerful but more maneuverable air superiority systems, thereby misrepresenting the real capability of air superiority aircraft. This explains the "curious" location of the F-14 on Snider's graphical display of aircraft capability (Figure 3, p. 60).[39] According to Snider's analysis, the F-14A is inferior to the MiG-23, MiG-25, F-4E, and F-15A as an air superiority platform. In reality, it is inferior only in terms of engine thrust and speed, not in

---

[39] Snider's results would better sort themselves out if instead of a single dimension for interceptors and air superiority systems, separate indices were constructed for each aircraft type.

FIGURE 3

CAPABILITY MIX OF PRINCIPAL COMBAT AIRCRAFT TRANSFERRED TO THE THIRD WORLD, 1945-1973



terms of maneuverability, and certainly not in terms of its weapons suit. The dangers and pitfalls of oversimplification are apparent in this particular instance.

Some people argue that weapons capability is much more than a combination of performance characteristics. With respect to aerial combat capability, for example, there is empirical evidence which suggests that pilot/crew proficiency is the ultimate determinant of fighter capability. The Navy's fighter community can trace much of its combat success during the latter stages of the Vietnam War to better training and crew proficiency. Prior to 1968, the Navy's kill ratio against the North Vietnamese was a rather unimpressive 2.9 to 1. After the establishment of the Fighter Weapons School, the figures steadily improved until, by 1972, it had reached 12 to 1.[40] Another pertinent example is found in the Middle East where the superior training and tactics of Israeli pilots have paid huge dividends in maintaining air superiority against countries equipped with some of the best aircraft in the world.

The importance of operator proficiency can be generalized to other weapons and combat milieu. Quoting Kemp:

> ...the superiority of Israel's armed forces in terms of their ability to use weapons effectively has meant in the past that the qualitative difference between Egypt's modern T-55 tanks and Israel's less modern Centurions has not been so important as would have been the case if both sides were equally proficient at fighting with these weapons.[41]

[40] Aviation Week and Space Technology, p. 62, 3 December 1973.

[41] Kemp, Geoffrey, Classification of Weapons Systems and Force Designs in Less Developed Country Environments C/70-3, p. 32, MIT, 1970.

Kemp's work on weapons classification and force design isolates other factors which impact on weapon system potential. Mainly, he argues that the only realistic way to assess a weapon or military force's potential quality is to set the technological qualities of the system or force against the unique environmental factors of the operating area. Of the six specific environmental characteristics Kemp feels are significant,[42] three in particular can be linked to capability assessment:

(1) the geography of the area

(2) the combat environment, that is, the technical competence, fighting ability, and force levels or the adversary

(3) the technical competence and fighting ability of the force using the weapon.

Geography is most applicable to ground system capability assessment since air and naval systems have stable operating environments world-wide. One example showing geographic effects on ground system capability is provided by Snider who points out that the intrinsic amphibious capability of Soviet tanks is tremendously important to the Egyptians because of the Suez Canal but is hardly relevant in countries

---

[42] Kemp, Geoffrey, "Arms Traffic and Third World Conflicts," International Conciliation, No. 577, p. 25, March 1970.

like Algeria and Afghanistan.[43] A more graphic example is provided by the Insight Team of the London Sunday Times. As Figure 4 depicts,[44] Soviet tanks exhibit low profiles and little gun depression capability, and as such, are optimized for flat terrain. In contrast, U.S. and British tanks have larger turrets and greater gun depression capability which make them better suited for the undulating terrain of the Middle East. When geography optimizes or degrades a weapon's performance in this fashion, corresponding adjustments in the weapon's assessed capability must be made in order to bring a certain degree of reality to the assessment.

According to Kemp's findings, the utility of any weapon system is determined largely by the combat environment. As Tables V through VII (pp. 65-67) show, different aircraft are preferred under different circumstances because the mission requirements vary. Observe, for instance, that high speed capability is important in a hostile environment but becomes unnecessary in permissive situations. Even within the hostile environments categorized, the relative importance of speed varies. Inasmuch as utility is a measure of

---

[43] Snider, Lewis W., Arabesque: Untangling the Patterns of Supply of Conventional Arms to Israel and the Arab States and Their Effects on the Arab/Israeli Conflict 1948-1973, paper presented at the 1975 Annual Convention of the International Studies Association, Washington, D.C., 19-22 February 1975, p. 10.

[44] Insight Team of the London Sunday Times. The Yom Kippur War. Garden City, N.Y., Doubleday, 1974.

FIGURE 4

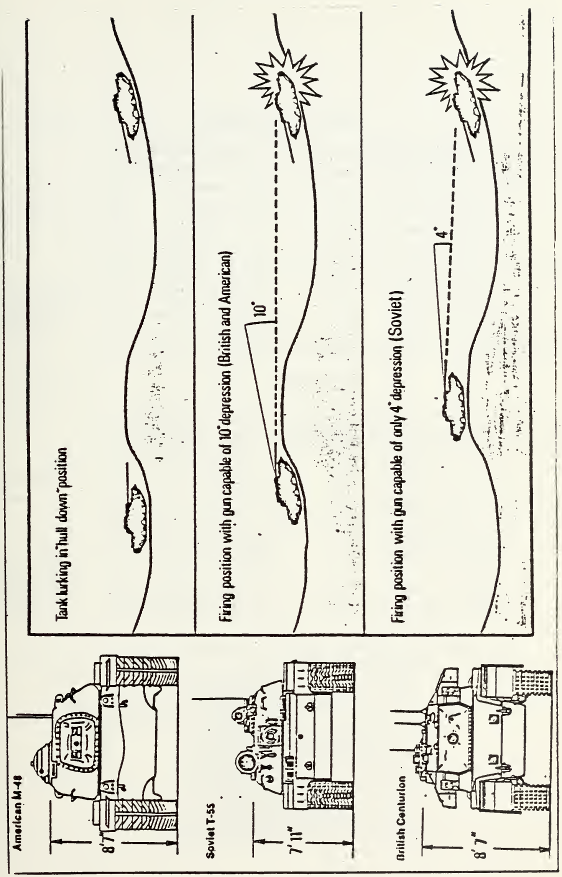

American M-48
8'7"
Soviet T-55
7'11"
British Centurion
8'7"
Tank lurking in "hull down" position
Firing position with gun capable of 10° depression (British and American)
10°
Firing position with gun capable of only 4° depression (Soviet)
4°

TABLE V

PERMISSIVE COIN

	ISOLATED AIR STRIKE		TACTICAL SUPPORT	
	Mission 1		Mission 2	
Preferred Systems	1st	2 Squadrons B-26K	1st	1 1/2 Squadrons B-57
	2nd	1 1/2 Squadrons B-57	2nd	1 Squadron A-37B
	3rd	1 Squadron A-4F	3rd	2 Squadrons B-26K
	4th	2 Squadrons A-37B	4th	1 Squadron A-4F
Characteristics Influencing Selection of 1st Choice	1st	High Total Payload	1st	High Total Payload
	2nd	Multiple Armament	2nd	Multiple Armament
	3rd	Good Loiter Capability	3rd	Good Loiter Capability
	4th	Combat Well Tested	4th	Good Low Altitude Performance

HOSTILE COIN

	ISOLATED AIR STRIKE		TACTICAL SUPPORT	
	Mission 5		Mission 6	
Preferred Systems	1st	1 Squadron A-4F	1st	1 Squadron A-4F
	2nd	1/2 Detachment F-4C	2nd	1 1/2 Squadrons B-57
	3rd	1 1/2 Squadrons B-57	3rd	2 Squadrons A-37B
	4th	1 Squadron F-5A	4th	1/2 Detachment F-4C
Characteristics Influencing Selection of 1st Choice	1st	High Total Payload	1st	High Total Payload
	2nd	High Speed Capability	2nd	Multiple Armament
	3rd	Multiple Armament	3rd	Good Loiter Capability
	4th	High Combat Radius	4th	High Speed Capability

Package (1 Squadron = approx. 12 aircraft;
1 Detachment = approximately 6 aircraft)

Choices of Systems Offered for All Attack Missions
2 SQUADRONS A-37B
1 SQUADRON F-5A
1 1/2 SQUADRONS CANBERRA (B-57)
1 SQUADRON A-4F
1 SQUADRON SU-7
2 SQUADRONS OV-10A (BRONCO)
1 SQUADRON MIG-21D
2 SQUADRONS B-26K
1 1/2 SQUADRONS HUNTER OR F-86
1 DETACHMENT LIGHTNING
1 DETACHMENT MIRAGE III
1/2 DETACHMENT F-4C

TABLE VI

PERMISSIVE EXTERNAL

		ISOLATED AIR STRIKE MISSION 3		TAC SUPPORT MISSION 4
Preferred Systems	1st	1 1/2 Sqaadrons B-57	1st	1 1/2 Squadrons B-57
	2nd	2 Squadrons B-26K	2nd	1 Squadron A-4F
	3rd	1 Squadron A-4F	3rd	2 Squadrons B-26K
	4th	2 Squadrons A-37B	4th	2 Squadrons A-37B
Characteristics Influencing Selection of 1st Choice	1st	High Total Payload	1st	High Total Payload
	2nd	Multiple Armament	2nd	Multiple Armament
	3rd	Good Loiter Capability	3rd	Fast-turn Round Time
	4th	Good Low Altitude Performance Ease of Maintenance	4th	Good Low Altitude Performance

HOSTILE EXTERNAL

		ISOLATED AIR STRIKE MISSION 7		TAC SUPPORT MISSION 8
Preferred Systems	1st	1 Squadron A-4 F	1st	1 Squadron A-4F
	2nd	1/2 Detachment F-4C	2nd	1/2 Detachment F-4C
	3rd	1 Detachment Mirage III	3rd	1 Detachment Mirage III
	4th	1 Squadron F-5A	4th	Squadron F-5A
Characteristics Influencing Selection of 1st Choice	1st	High Total Paylod	1st	High Total Payload
	2nd	Multiple Armament	2nd	Multiple Armament
	3rd	High Speed Capability	3rd	Good Loiter Capability
	4th	Good Low Altitude Performance	4th	High Speed Capability

TABLE VII

DEFEND AIRFIELDS MISSION 9	DEFEND CITIES MISSION 10
1. 1 Squadron MIG-21D	1. 1 Squadron MIG-21D
2. 1 Detachment Lightning	2. 1 Squadron F-5A
3. 1/2 Detachment F-4C	3. 1/2 Detachment F-4C
4. 1 Squadron F-5A	4. 1 Detachment Lightning
1. High Speed Capability All -Weather Capability	1. High Speed Capability All-Weather Capability
3. Short Runway Requirements	3. High Total Payload
4. High Total Payload	4. Multiple Armament

CHOICES OF SYSTEMS OFFERED FOR AIR DEFENSE MISSIONS

PACKAGE (1 Squadron = Approx. 12 aircraft 1 Detachment = Approx. 6 aircraft)
1 SQUADRON F-5A
1 SQUADRON MIG-21D
1 1/2 SQUADRONS HUNTER/F-86
1 DETACHMENT LIGHTNING
1 DETACHMENT MIRAGE III
1 SQUADRON F-100D
1 SQUADRON MIG-19
1/2 DETACHMENT F-4C

capability, Kemp's preference data implies that differences in performance characteristics need not equate to any practical differences in capability. This point is substantiated by the clear preference for older, less sophisticated systems, such as the A-4F and the B-26K.

A final important external variable affecting weapons capability is the technological capacity of the owner. It does little good to have a sophisticated avionics/weapons package in an aircraft if it cannot be kept in operating condition. Such a concern is minimal when assessing the capability of weapons owned and operated by militarily developed nations (e.g., Israel), but for African or Latin American equipment, technical capacity would certainly affect capability levels, most probably in a negative way. The chances of a Russian-maintained MiG-21 having all of its systems in operating condition are much better than for an Egyptian-maintained MiG-21 because the Egyptians face uncertainties in parts supply, and their maintenance skills are less developed. This corresponds to arguments presented by Snider and Kemp that simple, more durable systems have more military utility in Third World environments than sophisticated ones.[45]

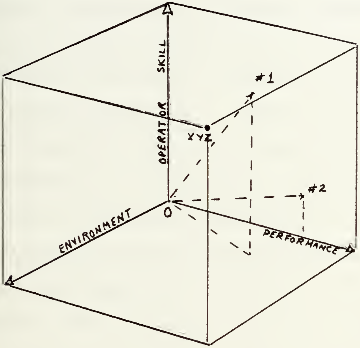
To summarize, defining capability using what has been termed the multi-faceted approach involves integrating the weapon's performance characteristics with operator skills,

---

[45]See in particular Kemp (1970) and Snider (1975).

and environmental and technological influences. The significance of this integration is diagrammed in Figure 5 wherein capability is represented as a vector in a multi-dimensional space.

FIGURE 5



The origin, 0, is a theoretical reference point that connotes no capability at all, while position X,Y,Z indicates maximum system capability. On this basis, system #1 has more total capability than system #2 because higher pilot skills and a more favorable technical environment outweigh system #2's built-in performance edge. Such an analysis is not possible with the weapon's performance approach since system #2's performance advantage would always translate to higher capability relative to system #1.

Having proposed two conceptual definitions for aerial combat capability — the first with weapons performance characteristics exclusively, and the second using information on the operator, environment, and performance capacity — it becomes important to present the strengths and weaknesses of each as they pertain to measurement and scaling. Initially the weapons performance approach appears to offer fewer measurement problems because all of the variables can be expressed in natural or derived units of physical measure, e.g., length, width, etc. However, two significant uncertainties must be faced in the operationalization process: (1) determining if weighting the variables in some fashion would lead to a better representation of capability; and (2) determining how the variables should be combined (i.e., added, multiplied, etc.). Little or no effort has been made to clarify these issues in connection with capability assessment. Hence, while using weapon's performance characteristics facilitates measurement because they are amenable to quantification, the characteristics themselves provide little guidance as to how a capability scale should be constructed.[46]

---

[46]Solutions to both of these questions are embodied in the process leading to factor scores where weighting is determined by the way a variable loads on a factor, and addition is rigidly assumed. (These patented solutions may partially explain the use of factor analysis by Muhalka and Snider.) However, there is nothing sacred about factor analysis or factor scores. The assumption that linearity holds may be inappropriate for combining variables within a factor to describe capability. Moreover, if more than one factor is needed to describe capability, the correct combination <u>among</u> factors becomes a problem. Questions can also be raised about the construct validity of factors. The fact that certain variables cluster together may, in actuality, not contribute to capability assessment at all.

Using the weapons performance approach may also give rise to content validity problems to the extent that key variables may be omitted. To reinforce a point made earlier, the analyst must cope with this issue through technical research on the weapon type. However, even with such an effort, there are no assurances that meaningful information will not be overlooked.

Reliability problems also plague the performance approach, first because of the disparities in open source estimates of the performance variables, and second because of the nature of the variables themselves. During the course of this research, for instance, substantive differences in aircraft performance characteristics were commonplace even for the most straightforward items such as velocity and range. Without any <u>difinitive</u> sources, the chance for error — usually in the form of an over-estimate in capability — is high. Furthermore, the dynamic nature of the variables accentuates the reliability problem. Consider, for a moment, the thrust-to-weight ratio (T/W). In some sources it is calculated using the normal take-off weight of the aircraft while in others it is derived using the combat weight (i.e., the basic system weight with one-half internal fuel and weapons). Since the difference between these two weight values can be as much as 10-20 per cent of the total system weight, T/W values vary tremendously. In most cases, the researcher has no way of knowing which value was used. The same can be said of wing loading, W/S. Hence, the danger of misrepresenting system capability using these variables is difficult to escape.

With respect to the multi-faceted approach, it is felt that using it to scale capability can potentially yield a more accurate representation of true system capability and better accommodate its dynamic nature. However, the analyst is again faced with the problem of determining the relative importance of variables. The vector space illustrated in Figure 5 (p. 69) assumes that weapon performance characteristics, operator proficiency, and the environment all affect capability equally. In reality, this may not be the case.

One important deficiency found in the multi-faceted approach is a genuine difficulty in operationalizing each variable. While it may be possible to devise an algorithm for operator proficiency based, for example, on flight time or training, there are no clear procedures for such attempts and potential validity problems. The same can be said for operationalizing technological capacity or the environment. In the final analysis, these are judgemental variables with no obvious physical correlates and as such will always be difficult to quantify.

Given the concerns over variable selection/omission, weighting, operationalization, and the like, using a strict analytic procedure to derive capability scores may not be the best available approach. Judgemental scaling techniques which could tap the knowledge of weapons experts, systems operators, etc., may be more effective since the subtleties of the analysis are preserved and resolved in the judgemental process. This proposal will be explored in Chapter IV.

IV. POSSIBLE METHODS FOR SCALING CAPABILITY

In Chapter II some important reasons for developing capability indices were mentioned. Among the most important was that such measurement would facilitate side-by-side comparisons with equipment from other countries and provide guidance to military and political decision makers concerning the most appropriate weapons to transfer in order to preserve or alter the military balance in an area. Chapter III presented two conceptual approaches to capability and isolated important variables related specifically to aerial combat capability. Additionally, the complexity of capability measurement was stressed and the problems of measuring capability discussed. This chapter embodies the next logical step by presenting four scaling techniques which can be used to measure capability —factor analysis, paired comparisons, successive intervals, and multi-attribute utility analysis. Each method is discussed separately with emphasis placed on the rationale for using the method, and its theoretical basis. Each method is also applied to the problem of scaling aerial combat capability and the results of the applications are discussed.

An important theme running through this chapter is that the three judgemental scaling techniques — paired comparisons, successive intervals, and multi-attribute utility scaling — represent viable options to factor analysis for capability assessment. In fact, it is the author's contention that they presently offer the best solutions to measuring capability. There are three fundamental reasons for this view. First, since judgemental techniques depend on expert evaluations, it can be assumed that the resulting scores reflect the best synthesis of all relevant information. Second, the analyst is not burdened with the task of selecting variables and specifying their exact interrelationship. This is done automatically in the judgemental process. Finally, using the judgemental technique allows the analyst to consider more than just performance characteristics. Environmental factors, tactics, and subtle variables such as pilot proficiency, crew coordination, cockpit visibility, and so on, can be considered without any added burden to the analyst. The reader is urged to keep these advantages in mind.

A. THE FACTOR ANALYTIC APPROACH

1. Rationale

Despite periodic references to some of the pitfalls encountered when trying to scale capability using factor analysis, there are many compelling reasons for exploring the technique in more detail.

First, as a general scientific method for analyzing data, factor analysis has been used with apparent confidence by scholars from many disciplines. Study and theorizing in economics, sociology, anthropology, biology, and political science, all have been facilitated through a use of factor analysis.[47] The breadth of usage alone is an enticement to examine the approach fully before passing judgement on it.

Second, it is compatible with SPSS (Statistical Package for the Social Sciences), and as such can be used easily and efficiently by the research community. Inasmuch as one of the prevalent burdens on current research is touching base with vast amounts of data and information, having the option of using an existing computer program is most attractive.

A third, and more pragmatic, reason for exploring factor analysis is to derive a set of aircraft capability scores using the variables advanced in Chapter III in order to more fully examine the reasonableness of Snider's and Mihalka's results. Hopefully, such a comparison will be heuristic and provide valuable insight into the strengths and weaknesses of the method.

2. Theory

As a general methodology, factor analysis can be summarized with three key concepts:[48]

(1) the concept of patterned variation

(2) the concept of vector spaces

(3) the concept of dimensionality.

---

[47]Rummel, op. cit., p. 13.

[48]Ibid., pp. 13-19.

Discovering patterns and uniformities in the variation of data is significant because it is one indication that meaningful relationships exist. Factor analysis searches for such variation in one of two possible ways: by scanning the cases of a data matrix to see if any exhibit similar characteristics (Q-analysis); or by searching through the characteristics themselves for regularities and patterns (R-analysis). R-analysis has greater pertinence to weapons capability assessment since the only thing Q-analysis would accomplish would be a categorization of weapons by mission-type. Table VIII, which presents weapons performance characteristics for four aircraft, illustrates this point.

TABLE VIII

	MSPEED	CEILING	COMBAT RADIUS	THRUST-TO WEIGHT	WING LOADING
F-15A	M 2.5	60 K ft	470	1.25	55
F-14A	2.2	58 K	480	0.86	40
A6E	0.63	45 K	800	0.35	100
A7D	0.67	45 K	750	0.40	94

Profile or Q-analysis delineates two groups — an air-superiority group characterized by relatively high speeds, ceilings, and thrust-to-weight ratios, and low wing-loadings and combat radii; and an attack grouping described by lower speeds, ceilings and thrust-to-weight ratios, and higher combat radii and wing loading. These similarities are apparent without factor analysis. What is less obvious, yet more important for determining capability, is the relationships among the characteristics. R-analysis addresses this issue.

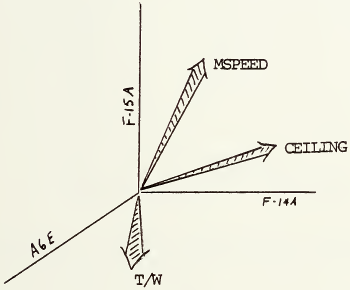
Knowledge of vector algebra and vector spaces can enhance the understanding of factor analysis by providing visual meaning to the data and the factoring process. The data matrix in Table VIII, for instance, can be viewed as a series of row and column vectors, each having a certain magnitude and spatial direction. The magnitude depends on the values of the individual elements composing the vector, e.g., MSPEED = (2.5, 2.2, .63, .67), while the direction depends on the particular relationship that variable has with those remaining. Imagine, for illustrative purposes, that the F-15A, F-14A, and A-6E can describe the three-dimensional space given in Figure 5. Each of the five characteristics in Table VIII could be plotted as vectors in this space. The resulting angles between these vectors would measure the relationships among them for the three aircraft describing the space.[49] Small angular differences approaching zero degrees indicate a strong relationship, while angles near 90 degrees suggest no correlation at all.

---

[49]The basis for this discussion is taken from Rummel, R.J., "Understanding Factor Analysis," The Journal of Conflict Resolution, Vol. XI, No. 4, December 1967.

In the more extreme cases, obtuse angles connote a negative relationship, while a 180 degree difference implies that the vectors are inversely related. Thus, in Figure 5A. speed and service ceiling are strongly related while thrust-to-weight performance is much less so.

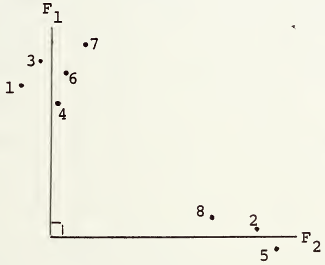
FIGURE 5A



Obviously, a geometric analysis becomes impossible when more than three dimensions are involved. Patterns of variation are obscured, clusters of vectors are not clearly delineated, and relationships are not apparent. The way factor analysis deals with this brings up the notion of dimensionality. As noted above, vectors which cluster together are highly related. This being so, it is possible to represent an entire cluster of variables with one mathematically determined line — called a dimension or factor —

which projects through the cluster and defines its variance. In this fashion, many variables are reduced to a manageable number of factors which depict regularities within the data. Figure 6 displays a plot of eight variables which can be represented with two factors, F_1 and F_2.

FIGURE 6



The angular separation between factors is significant and should be discussed briefly. Notice in Figure 6 that the two factors describing the variable clusters are 90 degrees apart. The fact that these orthogonal axes successfully represent the clusters implies that the clusters are basically uncorrelated. If the analyst suspects orthogonality, or wishes to measure the degree of orthogonality in the data, orthogonal factoring (usually varimax)[50] is warranted.

However, circumstances arise when orthogonal factors do not provide the best representation of variable clusters.

---

[50]Rummel, op. cit., 1970, p. 392.

Figure 7-a, for example, shows two variable clusters located near each other between orthogonal axes. No matter how the orthogonal axes are rotated, they will never precisely delineate the relationship between the clusters. The general solution is to drop the demand for uncorrelated factors and use an <u>oblique</u> rotation scheme (Figure 7-b). However, as Rummel notes, controversy exists over the use of oblique rotation methods.[51] Some people feel that oblique rotation is a way to fabricate relationships between variables which, in reality, do not exist. Others, including Rummel, support oblique rotation on the grounds that it generates more precise information and relationships between clusters and that the process better reflects reality.

> ...the real world should not be treated as though phenomena coagulate in unrelated clusters. As phenomena can be interrelated in clusters, so the clusters themselves can be related. Oblique rotation allows this reality to be reflected in the loadings of their factors and their correlations.[52]

While there are obvious advantages gained through orthogonal factor rotation in terms of simplicity and amenability to mathematical manipulation, they do not justify dismissing oblique rotation in capability analysis, especially since capability analysis is an open-ended issue at this point in time. Both approaches should be used.

---

[51]Rummel, <u>op. cit.</u>, 1967, p. 477.

[52]<u>Ibid</u>.

FIGURE 7-a

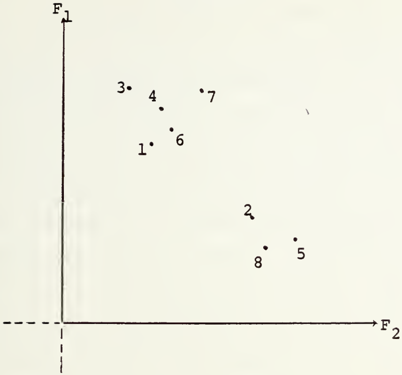
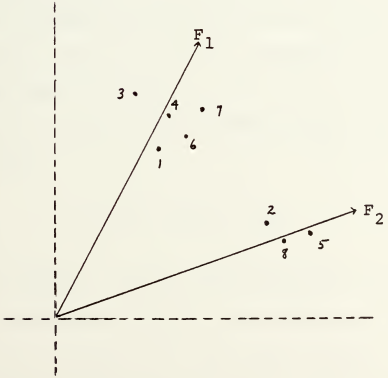


FIGURE 7-b



An understanding of the three principles of factor analysis — variability, vector representation, and dimensionality — provides the framework needed to understand factor analytic scaling and measurement. The basic mechanism for factor analytic scaling is to calculate factor scores in the following way. After the patterns in variation have been identified and factors described, each individual variable is weighted proportionally to its relationship with a given factor. The greater or stronger the relationship, the greater the weight. The value of that variable for a particular case is then multiplied by the weighting (loading) and normalized to yield a score. Since factors are usually a composite of many variables, i.e., a variable cluster, this weight-times-data product must be computed for every variable associated with the factor and summed to provide the actual factor score. Symbolically, if α_1, α_2, and α_3 represent the loadings of three variables on a factor, and if the values of these variables are v_1, v_2, and v_3 respectively, the factor score, S, would be:

$$S = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 \tag{9}$$

In short, a factor score measures the magnitudes of the variables for a given case and their relationships to a factor. If the factor has been properly identified, and the analyst is convinced that it legitimately represents —

either in a descriptive, causal, or symbolic sense — an attribute or phenomenon, the factor scores can be used to rate cases along the factor with interval-level precision.

3. Scaling Aerial Combat Capability With Factor Analysis

While the theory behind factor analysis gives credence to factor scores, Snider's and Mihalka's results (Figure 3 and Table II) raise two questions that must be resolved before factor analytic scores can be accepted as true representations of weapons capability. First, does the two-factor structure assumed by Snider and Mihalka sufficiently represent a weapon's capability, or is this structure a function of the variables used? Second, assuming there are multiple factors, is there any clear way to combine or weight them that will lead to a meaningful representation of the system's capability?

To answer the first question a data matrix of 29 aircraft (including both attack and air superiority systems) and 13 variables (Snider's plus some of those identified in Chapter III), was factor analyzed using both principal factor analysis and varimax rotation. The variables used were defined in the following way:

(1) Maximum speed - the fastest advertised speed of the system

(2) Service ceiling

(3) Thrust - total maximum engine thrust at sea-level. If the system had after-burner capability, the maximum A/B thrust value was used.

(4) Normal take-off weight - the weight of the fuel + platform + ordnance in its normal mission configuration.

(5) Rate-of-climb - measured in feet per minute at sea level.

(6) Maximum payload - the maximum amount of ordnance the system can carry.

(7) Combat range - the maximum one-way distance the aircraft can fly and perform its mission.

(8) Combat Radius - the maximum two-way distance the aircraft can fly and perform its mission.

(9) Thrust-to-weight ratio - ratio of maximum engine(s) thrust to normal take-off weight.

(10) Wing loading - ratio of normal take-off weight to total wing area.

(11) GUNB - the total number of gun barrels on the aircraft.

(12) MISALG - a crude algorithm representing missile capability based on the range of the system (at 10-15000 feet AGL), the number of missiles the aircraft normally carries, and an angular estimate of the firing sector.

(13) Production year - the year production began. If the aircraft was an updated model, the year the modification occurred was used with the thought of capturing the incorporated technological advances.

The resulting varimax factor matrix and computed factor scores are displayed in Tables IX and X. The first point to note is the presence of three distinct factors in contrast to the two which Snider and Mihalka obtained. Based on the particular variables clustering together, Factor I generally corresponds to Snider's Interceptor/ Air-to-Air Combat - Factor (see Table IV). Similarly, the same variables which load on Factor II also load together on Snider's Tactical Support/Ground Attack - Factor. To this extent, Snider's results have been replicated. However, all of the variables not considered by Snider — thrust-to-weight ratio, wing loading, GUNB, and MISALG — cluster together to form a third factor. Assuming that the analysis in Chapter III is correct and the added variables are important to air-to-air combat capability, the presence of the third factor confirms what has previously been suggested, namely, that measuring air-to-air combat capability with Factor I alone is inadequate.

Some additional observations should be mentioned. Production year, which loaded highly on the Air-to-Air Combat - Factor for Snider, loaded on the new factor. This is not necessarily inconsistent since it is argued that both Factor I and Factor III relate to the same attribute. Also

TABLE IX

VARIMAX ROTATED FACTOR MATRIX
FOR 29 AIRCRAFT AND 13 VARIABLES

VARIABLE	FACTOR I	FACTOR II	FACTOR III
Max Speed	.91183	-.16005	.15425
Ceiling	.90017	-.14516	-.10637
Thrust	.81375	.33873	.27959
Rate-of-climb	.85771	-.17088	.31725
Take-off weight	.62739	.68222	-.04521
Payload (max)	-.22243	.91291	.07798
Combat range	-.06186	.90778	.01947
Combat radius	-.09686	.90804	.00532
Thrust-to-weight	.54453	-.32122	.54158
Wing loading	.07857	.34959	-.83717
Number of gun barrels	.07818	.13349	.88188
Missile algorithm	.30709	.24849	.52984
Production Year	.27103	.40090	.52844

of interest is the fact that the thrust-to-weight ratio loeads equally on Factors I and III. This seems to reinforce the point just made that both Factors I and III are needed to represent air-to-air combat capability inasmuch as T/W is critical to an air-superiority platform. It also implies that not all variables will be clearly associated with any one factor in particular, at least when employing varimax rotation. This represents a potential problem from the point of view of defining and interpreting the factors.

The impact of the added variables and the multiple-factor result is more apparent in the Factor Score Matrix (Table X) and the derived system rankings in Table XI. Depending on the factor — or combination of factors — chosen, the ranking of aircraft varies. The problem is obvious. How must the factors be weighted and combined to produce the best results? With respect to fighter aircraft, the author tends to favor the combination of Factors I and III because the variable clusters included are more important to the aerial combat mission than are the variables in Factor II. However, this cannot be considered a definitive equation for aerial combat capability since the question of weighting factors has yet to be resolved. Should Factor I and Factor III be viewed equally, or must Factor III be considered more heavily than Factor I to generate true capability scores?

In an attempt to focus more closely on air-to-air combat capability and fighter aircraft, the variable list was reduced from thirteen to eight, all of which came from

Table X

Factor Scores for 29 Aircraft
13 Variables Varimax Rotation

Aircraft	Factor I	Factor II	Factor III
A-37B	-1.6193	- .8519	- .2089
A-4n	-1.3927	.2733	.0683
A-4E	-1.5728	.3169	.0236
A-4F	-1.5728	.3169	.0236
A-6E	-0.7542	3.1342	-1.2066
A-7D	-1.0957	2.0996	-0.8608
Su-7B	-0.3517	- .5716	- .3454
Su-7MF	.1695	- .6426	- .3271
Hunter Mk6	-1.3381	- .6550	- .0229
Lightning Mk2	.0471	- .8567	- .1541
Lightning Mk53	.6861	.6190	- .7096
Su-9	.4508	- .5456	- .6531
Su-11	.5492	- .8610	- .7768
Faithless	.6350	- .9714	- .8727
Mirage 3C	.1024	- .4651	.1452
Mirage 3E	- .1252	- .3499	- .0099
Mirage 5	.0938	- .0942	.0117
MiG-19	- .5383	- .9020	- .2701
MiG-21PF	.0428	-1.1306	- .2305
MiG-21MF	.3717	-1.2125	- .2252
MiG-23	.3456	- .2657	.4424
MiG-25	3.0011	0.5937	-2.1415
F-5E	- .3523	- .3733	.0570
F-4B	.5039	.7131	- .0703
F-4E	.7385	.6729	- .0773
F-14A	.6848	.9899	1.6640
F-14A (with Phoenix)	.7986	1.3007	2.6069
F-15A	1.3223	.2798	1.7756
F-16	.1699	- .5719	2.3445

Table XI

Rank Order for 12 Fighter Aircraft Using Different Factor Combinations

Factor I alone	Factor I + Factor III	All Factors
MiG-25	F-15A	F-15A
F-15A	F-14A (with Phoenix)	F-14A (with Phoenix)
F-4E	F-16	F-14A
F-14A (with Phoenix)	F-14A	F-16
F-14A	MiG-25	MiG-25
MiG-23	MiG-23	F-4E
MiG-21MF	F-4E	MiG-23
F-16	Mirage 3C	Mirage 3C
Mirage 3C	MiG-21MF	F-5E
MiG-21PF	MiG-21PF	MiG-21MF
F-5E	F-5E	MiG-21PF
MiG-19	MiG-19	MiG-19

Chapter III. Additionally, the number of aircraft was reduced to provide a more homogeneous set of 21 interceptor/ air superiority platforms. These changes also serve to test the three-factor structure and the various rank-orderings obtained previously.

The factor matrices for the varimax and oblique rotations of the eight-variable data set are presented in Tables XII and XIII. While the basic variable patterns are similar to those obtained in the first factor analysis, a number of changes have occurred. First, the thrust-to-weight ratio now loads strongly on only one factor instead of moderately on two. Conversely, wing loading relates moderately to two factors in the latest analysis rather than one.[53] Shifts like this can be expected with changes in cases and variables, and their effect on capability assessment and the factor scores is noticeable. Table XIV shows the rank-ordering obtained from the second set of varimax factor scores. Comparing these rankings with the previous rankings in Table XI for corresponding factor combinations[54] shows that occasionally a system will shift

---

[53]All 29 aircraft were factor analyzed in terms of the eight variable as well. In this instance thrust-to-weight ratio loaded equally on two factors (Factors I and II) rather than on one factor. On the other hand, GUNB loaded on Factor II alone instead of the dual-loading exhibited in Table XII. This shows that the variable clusters depend on the types of aircraft analyzed. The complete results are contained in Appendix I.

[54]For comparisons with Table II, consider Factor I on Table XI and Table XII to be the same, and Factor III of Table XI as equivalent to Factor II on Table XII.

Table XII

Varimax Rotated Factor Matrix
For 21 Interceptor/Air Superiority Aircraft

Variable	Factor I	Factor II	Factor III
Max Speed	.8952	.1564	.1961
Ceiling	.9103	-.1347	-.0115
Combat Radius	.3237	.1509	.6607
Thrust-to-weight	.1068	.8973	-.2058
Wing Loading	.3256	-.5500	-.4816
GUNB	-.1836	.7031	.5484
MISALG	.0335	.0422	.8435
Production Year	.0944	.7286	.3567

Table XIII

(Oblique Rotation)
Factor Pattern Matrix For
21 Interceptor/Air Superiority Aircraft

Variable	Factor I	Factor II	Factor III
Max Speed	.1875	.8863	-.1754
Ceiling	-.0716	.9140	.0043
Combat Radius	.0532	.2947	-.6652
Thrust-to-weight	.9696	.1032	.3270
Wing Loading	-.4515	.3535	.4357
GUNB	.6065	-.2161	-.4846
MISALG	-.1138	-.0019	-.8717
Production Year	.6887	.0699	-.2792

Factor Correlations

	Factor I	Factor II	Factor III
Factor I	1.0000	-.0497	-.2945
Factor II	-.0497	1.0000	-.0321
Factor III	-.2945	-.0321	1.0000

Table XIV

Rank Order for 12 Fighter Aircraft Using Different Factor Combinations

Factor I alone	Factor II alone	Factors I and II	All Factors
MiG-25	F-16	F-15A	F-14A (with Phoenix)
F-15A	F-15A	F-16	F-15A
MiG-21MF	F-14A	MiG-25	MiG-25
F-14A	MiG-23	F-14A	F-14A
MiG-23	F-14A (with Phoenix)	MiG-23	F-16
Mirage 3C	MiG-21MF	MiG-21MF	MiG-23
F-4E	F-5E	F-14A (with Phoenix)	Mirage 3C
F-14A (with Phoenix)	MiG-21PF	MiG-21PF	F-4E
MiG-21PF	F-4E	F-4E	MiG-21MF
F-16	Mirage 3C	Mirage 3C	MiG-21PF
F-5E	MiG-25	F-5E	F-5E
MiG-19	MiG-19	MiG-19	MiG-19

Note: For Comparisons with Table XI, Factor I on this Table corresponds with Factor I on Table XI while Factor II corresponds to Factor III on Table XI.

as much as four positions in either direction. The oblique factor scores display even greater disparity. (See Appendix I.)

Just how significant these differences really are is an intriguing question. In many instances, fairly high rank correlation (as measured by Spearman's rho) exists betwen two corresponding rankings. On the other hand, the magnitudes of the differences are not as important as the fact that no compelling reasons can be offered for using one factor method over another.

In contrast to the easily interpreted clusters obtained when thirteen variables are used (Table IX), interpreting the variable clusters from the eight-variable analysis is more challenging. Factor III (Table XII), for example, belies exact interpretation since it unites two important yet conceptually different variables — combat radius and missile capacity. Another source of confusion is whether to treat Factors I and II as representing a weapon type (e.g. interceptor) or a weapon attribute (e.g. energy or maneuverability). Snider and Mihalka both use the factors to identify aircraft types. However, with three factors appearing, even when just one type of aircraft is considered, the attribute interpretation gains credence. At any rate, these uncertainties constitute additional stumbling blocks in capability analysis.

The preceding investigation of the factor analytic method suggests three things:

(1) When analyzing aircraft, the number of factors obtained depends on the variables selected. Thus, the assumption that two clear dimensions will completely describe a weapon's capability cannot be supported as a general condition.

(2) Occasionally, important variables will not load on any one factor clearly. This creates factor-interpretation problems and raises questions as to the validity of factor scores.

(3) The ranking/scoring of aircraft depends on the factor combinations used. Until something is resolved as to the proper weighting and combination of multiple-factor situations, factor score results must be viewed with caution.

B. THE METHOD OF PAIRED COMPARISONS

1. Rationale

A fair amount of work has been done by psychologists to develop methods which measure qualitative phenomena such as beauty, affects, and excellence, all of which are difficult to define because they have no simple physical correlates. In a very real sense, scaling capability presents the same problem as scaling attitudes. It was noted, for instance, that selecting variables and specifying their exact relationships was difficult in the weapons performance approach.

The same was true of the multi-faceted approach which had additional operationalization problems. Inasmuch as judgemental techniques depend on the expertise of the judges, the operationalization problems are eased since it is assumed that the appropriate information will be weighted and processed automatically in the judgemental process. Judgemental techniques may be particularly valuable to policy analysts who either do not have access to technical data, or for one reason or another, do not feel qualified to select variables by themselves.

Of the many judgemental scaling techniques available, the paired comparisons method is particularly advantageous for several reasons. First, it is simple to administer. Information can be gathered either by questionnaire or through direct, oral responses. Second, in comparison to other judgemental techniques, such as the constant sum method and the subjective estimate method, it asks little of the judges. All that is assumed is that a judge can rank a pair of cases according to which has the greater or lesser amount of a particular attribute. This equates to ordinal-level information. Finally, the method has a certain amount of elegance to the extent that the requested ordinal information is transformed to interval measurement by making some simple assumptions about a judge's behavior. Capability scores would thus have the same precision as a factor score and potentially be more accurate.

2. Theory

Central to understanding the paired comparisons method of scaling is the law of comparative judgement, a set of statistical equations derived by psychologist Louis L. Thurstone. Thurstone postulated that when a stimulus or instance is presented to an observer it gives rise to a "discriminal process" which has a value to the observer on a psychological continuum.[55] Because of fluctuations in the observer's perception/judgement, the same instance will not result in precisely the same value all the time. However, there is a definite regularity in the estimating process such that the frequency distribution of judged values for a given instance over a large number of trials has a normal distribution.[56] A graphic example of the theoretical distributions of discriminal processes for two instances, *i* and *j*, is shown in Figure 8. The mean of each distribution, S_i and S_j, is the scaled value of each instance, while the standard deviations, σ_i and σ_j, determine their variance.

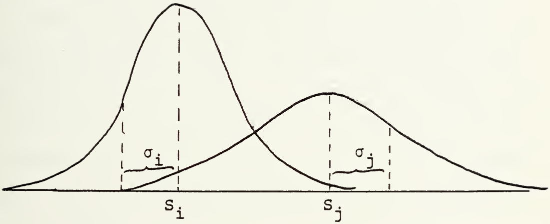
Since an individual's judgement of an instance or attribute has a normal distribution for a large number of trials, it follows that the difference between two instances presented simultaneously would also have a normal distribution. It turns out that the mean of such a distribution

---

[55] Torgerson, Warren S., Theory and Methods of Scaling, p. 159, John Wiley & Sons, 1967.

[56] Ibid., p. 159.

FIGURE 8



equals the difference between the means of the original distributions, $S_i - S_j$. Its standard deviation, σ_d, depends in turn on the standard deviations of the two original distributions (σ_i and σ_j), and on the correlation between them, r_{ij}. Mathematically, $\sigma_d = \sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j}$.

If instances _i_ and _j_ are judged a large number of times, the proportion of times one is rated higher than the other can be ascertained. For all practical purposes, this proportion can be viewed as the probability that the high-rated instance will continue to be rated superior in future trials. Treating the proportional results as probabilities is convenient because it permits the analyst to enter standard normal tables and convert the information to z-scores. In other words, if in Figure 8 instance _j_ is judged superior to instance *i*, the difference between the scale values, S_j and S_i, can be measured with a z-score value, z_{ij}, which relates to the proportion of times instance *j* was judged superior to instance *i*.

Since there is a certain amount of variance in the difference between S_j and S_i — measured in terms of the standard deviation, σ_d — the z_{ij} values will exhibit the same variance. Hence, the difference between S_j and S_i is more precisely written as $z_{ij}\sigma_d$. Substituting the complete expression for σ_d, the equation for the law of comparative judgment results.

$$S_j - S_i = z_{ij}\sqrt{\sigma_i^2 + \sigma_j^2 - 2\, r_{ij}\sigma_i\sigma_j} \quad (10)$$

where S_i and S_j are psychological scale values of the compared instances *i* and *j*, z_{ij} is the standard normal deviate associated with the number of times instance *j* was judged greater than instance *i*, σ_i is the standard deviation of stimulus i, σ_j is the standard deviation of instance *j*, and r_{ij} is the correlation between the deviations of the two instances.

As presented above, the equation describing the law of comparative judgment is not solvable since the number of unknowns exceeds the number of possible equations that can be generated. Hence, certain simplifying assumptions are normally made to arrive at a workable set of equations. Usually it is assumed that the standard deviation for the discriminal differences, σ_d, is constant.[57] This reduces equation (10) to

$$S_j - S_i = cz_{ij}$$

which, by allowing the constant, c, to equal 1, becomes,

$$S_j - S_i = z_{ij} \tag{11}$$

Although a simplified version of the law of comparative judgment, equation (11) can be used with confidence to derive scale values for the instances or objects that are judged. Thus, if there are n instances, n-1 equations of the same form as equation (11) will be obtained which, when solved simultaneously, give interval-level scores for every instance.

3. Scaling Aerial Combat Capability Using Paired Comparisons

To test the applicability of the paired comparisons method to capability analysis, a questionnaire[58] was distributed to 34 Navy and Air Force officers at the Naval Postgraduate School asking them to rate nine fighter aircraft according to the aircraft's aerial combat capability. The nine aircraft — MiG-19, MiG-21, MiG-23, Mirage-3C, F-4E, F5E, F-14A (without Phoenix), F-15A, and F-16 — were chosen

---

[57] Ibid., p. 165.

[58] A copy of the questionnaire is provided in Appendix II.

because of their current and/or potential use in Third World areas, and most especially in the Middle East. Care was taken to insure that every judge had a basic familiarity with every system by providing performance data on each aircraft. However, it was stressed that the judges were to rely on their personal knowledge and experience and not to base their decisions solely on the provided data.

To establish a reference scenario for the judges, it was stipulated that the combat environment was limited to altitudes below 20 thousand feet and speeds between M 0.5-1.5. These particular limits were based on open source documentation that the majority (i.e., 85-90 percent) of all aerial combat has occurred within these altitude and speed regimes.[59] In order to accentuate aerial combat capability, it was also stipulated that no long-range missiles such as Phoenix were available. This was also done to more accurately represent current Third World capability. Finally, all pilots were assumed to be of equal ability.

Since the most compelling reason for using a judgmental scaling technique in the desire to tap expert opinion and experience, the selection of judges is constrained by the particular system under consideration. For this reason 33 of the 34 officers who participated in the survey were aviators with fighter experience. The one exception was a

---

[59] Aviation Week and Space Technology, p. 41, 12 July 1971.

Navy pilot who, although not trained as a fighter pilot, possessed a unique knowledge of several of the Soviet systems gained through direct conversations with pilots who had defected from the Soviet Union.

The raw and transformed data gathered in the survey are presented in Tables XV, XVI and XVII. In Table XV, the raw frequency matrix, each individual block entry reflects the number of times the aircraft at the top of a particular column was judged superior to an aircraft at the beginning of a particular row. Thus, reading down the F-16-column, the 22 indicates that 22 of the 34 judges felt the F-16 was superior to the F-15A in terms of aerial combat capability. The corresponding blocks in Tables XVI and XVII represent the same information as a proportion and a z-score respectively.

Once the z-score array is calculated scale values are derived by taking column averages of the z-score values. Note, however, that the z-score matrix has a number of missing values. This occurs whenever one aircraft within a given pairing is judged superior to the other by every judge making the corresponding z-scores $\pm\infty$. Despite this problem, scale values can still be calculated by finding the average z-score difference between all pairs of z-scores in adjacent columns.[60] More precisely, if j and k are adjacent columns, and θ_{jk}

[60] This procedure is emphasized because it is felt that this problem will be a common one in weapons assessment.

Table XV

Raw Frequency Matrix
Paired Comparisons
of Nine Fighter Aircraft

	F-16	F15A	F14A	MiG 21	MiG 23	F4E	F5E	Mir-3C	MiG 19
F-16	-	12	1	0	2	0	0	0	0
F-15A	22	-	15	1	2	0	0	0	0
F-14A	33	19	-	0	1	0	0	2	0
MiG 21	34	33	34	-	17	13	13	14	9
MiG 23	32	32	33	17	-	12	13	17	12
F-4E	34	34	34	21	22	-	9	13	14
F5E	34	34	34	21	21	25	-	24	17
Mir 3C	34	34	32	20	17	21	10	-	9
MiG 19	34	34	34	25	22	20	17	25	-

Table XVI

Proportional Matrix

Paired Comparisons
of Nine Fighter Aircraft

	F-16	F-15A	F-14A	MiG 21	MiG 23	F-4E	F-5E	Mir 3C	MiG 19
F16	-	.353	.03	0	.059	0	0	0	0
F-15A	.647	-	.441	.03	.059	0	0	0	0
F-14A	.970	.559	-	0	.03	0	0	.059	0
MiG 21	1.0	.970	1	-	.5	.382	.382	.412	.265
MiG 23	.941	.941	.970	.5	-	.353	.382	.5	.353
F-4E	1	1	1	.618	.647	-	.265	.382	.412
F-5E	1	1	1	.618	.618	.735	-	.706	.5
Mir 3C	1	1	.941	.588	.5	.618	.294	-	.265
MiG 19	1	1	1	.735	.647	.588	.5	.735	-

Table XVII

Z-Value Matrix

Paired Comparisons
of Nine Fighter Aircraft

	F-16	F-15A	F-14A	MiG 21	MiG 23	F-4E	F-5E	Mir 3C	MiG 19
F-16	-	-.377	-1.88	-	-1.563	-	-	-	-
F-15A	.377	-	- .148	-1.88	-1.563	-	-	-	-
F-14A	1.88	.148	-	-	-1.88	-	-	-1.352	-
MiG 21	-	1.88	-	-	0	-.3	-.3	-.222	-.628
MiG 23	1.563	1.563	1.88	0	-	-.377	-.3	0	-.377
F-4E	-	-	-	.3	.377	-	-.628	-.3	-.222
F-5E	-	-	-	.3	.3	.628	-	.542	0
Mir 3C	-	-	1.352	.222	0	.3	-.542	-	-.628
MiG 19	-	-	-	.628	.377	.222	0	.628	-

is the set of rows having entries in both column j and column k, then the average scale difference $S_k - S_j$ is

$$S_k - S_j = \frac{\sum_{i \varepsilon \theta_{jk}} (z_{ik} - z_{ij})}{n_{jk}} \qquad (12)$$

With this format, eight equations with nine unknowns result.[61] However, since the scale produced is interval, the origin can be set arbitrarily. In this instance assigning a value to any one of the aircraft accomplishes the same thing. Thus, using the derived equations (Appendix II), and arbitrarily giving the MiG-19 a score of 1.0, leads to the following ranking and scale values:

Aircraft	Score
F-16	4.67
F-15A	3.80
F-14A	3.21
MiG-21	1.63
Mirage 3C	1.46
MiG-23	1.44
F-4E	1.27
F-5E	1.02
MiG-19	1.00

In terms of the rankings obtained with factor analysis, this arrangement correlates most closely with the ranking based solely on Factor II in Table XIV. (Spearman's rho = .97.)

---

[61] Appendix II contains the entire list of equations.

A review of Table XII shows that thrust-to-weight ratio, wing loading, GUNB, and production year, all load highly on this particular factor. On this basis, it is reasonable to conclude that some, if not all, of these particular factors were weighted more heavily than others in the judgmental process.

Because of the small sample size, no claim can be made that these particular paired comparison results actually reflect the preferences of the Navy and Air Force Fighter communities. It is possible that changes in the ranking would occur if more judges were polled. However, based on the large interval between the F-16/15A/14A and the other systems, it is safe to suggest that none of them would be ranked below any of the remaining systems and further that any changes that did occur would be among the lowest six.

In sum, while it is important to realize that the level of measurement is not improved, paired comparisons does represent a viable option to factor analysis.

C. THE METHOD OF SUCCESSIVE INTERVALS

1. Rationale

The method of successive intervals is another judgmental scaling technique that may have a place in capability analysis. Aside from the general advantage of relying on knowledgeable judges to analyze capability, which it shares with the paired comparisons method, the most immediate reason for demonstrating the method at this time is to obtain a set of results to compare with those obtained with paired comparisons. Consistent results, while not necessarily verifying either judgmental technique, will at least provide _some_ indication of their reliability.

The method is also worth considering because it provides information in a format which may be more useful than paired comparison results in certain instances. Specifically, it scales category boundaries, e.g., fair, good, excellent, on the same interval scale as the objects themselves. Thus, the analyst obtains information on the judges' estimates of the _meaning_ of a particular score, and not just the score itself.

2. _Theory_

Like the method of paired comparisons, the method of successive intervals relies on several theoretical assumptions about the judges' behavior. These assumptions are:

(1) that the psychological continuum a judge uses to assess the values of instances can be divided into a series of successively ordered categories

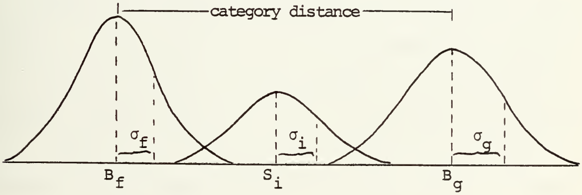
(2) that due to natural fluctuations in the judge's perception the category boundaries are not located in a fixed point but, rather, project a normal distribution of positions on the continuum

(3) that the judge will place a stimulus or instance below any category boundary whenever

the value of the stimulus/instance on the continuum is less than the category boundary.[62]

Taken together, these assumptions mean that the boundaries between categories behave like stimuli. A graphic interpretation of the theoretical distributions for an instance/object, S_i, and two boundaries, B_f and B_g, is found in Figure 9.

FIGURE 9



As noted in the previous discussion on comparative judgement, the difference between two normally distributed values will also project a normal distribution. Consequently, if B_g represents the mean value of the upper boundary, and S_i is the mean value of the instance, $B_g - S_i$ will be the mean value of the distance between the boundary and instance, and $\sigma_d = \sqrt{\sigma_i^2 + \sigma_g^2 - 2r_{ig}\sigma_i\sigma_g}$, its standard deviation.

---

[62] Ibid., p. 206.

Notice that this is the same expression encountered in the derivation of the law of comparative judgement with boundary information substituted for the second instance. After many judgements of the distance between B_g and S_i, a z-score can be obtained reflecting the proportion of times the instance was placed below the category boundary. Following the same steps which led to the law of comparative judgement provides the expression for the law of categorical judgement as well:

$$B_g - S_i = z_{ig} \sqrt{\sigma_i^2 + \sigma_g^2 - 2r_{ig}\sigma_i\sigma_g} \qquad (13)$$

Again, simplifying assumptions must be made to obtain solutions. In most cases, it is not unreasonable to assume that the instance-value and the boundary are stochastically independent random variables with a correlation coefficient, $r_{ig} = 0$. Moreover, it can also be generally assumed that all bounds have the same variance, so that $\sigma_g^2 = c$. (Other assumptions can be made depending on the circumstances and data. It is important to remember that they dictate the data analysis procedures required to obtain scale values).[63] Applying the previous assumptions, the simplified form of the law of categorical judgement becomes:

$$B_g - S_i = z_{ig} \sqrt{\sigma_i^2 + c} \qquad (14)$$

---

[63] Ibid., p. 209.

Given n instances and m category upper boundaries, mn equations in the same form as equation (14) can be derived. Because no assumptions are made about the variance of S_i, solving this set of equations is more difficult than is the paired comparisons procedure. The interested reader is referred to Appendix III for the recommended solution.

3. Scaling Aerial Combat Capability Using Successive Intervals

The same sample of Navy and Air Force pilots used in the paired comparisons scaling process participated in the successive interval scaling experiment as well. The aircraft scaled, combat parameters, and environment also were held constant. However, instead of pairing one aircraft against another, the judges were asked to categorize the aircraft as being either excellent, above average, average, below average, or poor with respect to its aerial combat capability. Appendix III contains a copy of the questionnaire used.

The raw frequency matrix is presented in Table XVIII. With the exception of one aircraft (the MiG-19), the judgments of each system clustered within two to three categories. The fact that the judgmental distributions are not spread over a wider range of categories indicates a certain level of agreement among the judges. However, it makes the derivation of capability scores more troublesome. In fact, interval scores cannot be calculated for the F-14A, F-15A, and F-16 because the method requires a frequency spread of at least three categories. This is a distinct weakness in

Table XVIII

Raw Frequency Matrix For
Successive Interval Judgments

	Poor	Below Ave.	Average	Above Ave.	Excellent
MiG-19	1	14	15	3	1
MiG-21	0	3	15	16	0
MiG-23	0	9	11	14	0
F-4E	0	9	16	9	0
F-5E	0	6	18	15	0
Mirage	0	3	16	15	0
F-14A	0	0	0	18	16
F-15A	0	0	0	9	25
F-16	0	0	0	1	33

the method. (Notice, however, that the F-14A, F-15A, and F-16 can be rank-ordered according to their distributions.)

One further concession is required. Since a raw frequency of zero leads to a $-\infty$ z-score it is necessary to compress the five original categories down to three (below average, average, and above average) in view of the fact that all but one of the remaining six aircraft exhibit this distribution in the first and last columns. It is important to stress that this reduction does not affect the score of any aircraft except the MiG-19, and that to an insignificant degree. Eliminating categories is simply an adjustment to the fact that "poor" and "excellent" are not pertinent to the aircraft listed.

After eliminating the F-14A, 15A, and 16, and reducing the number of columns, the frequency matrix appears in Table XIX. Converting the raw frequencies to z-scores and using the mathematical procedure outlined in Appendix III, leads to the following rank order and scores:

Aircraft	Score
F-16	No interval score possible.
F-15A	Ranking is based on original
F-14A	raw frequency data
MiG-21	.401
Mirage 3C	.338
MiG-23	.124
F-5E	-.023
F-4E	-.188
MiG-19	-.694

Table XIX

Adjusted Raw Frequency Matrix
For Successive Interval
Technique

	Below Ave	Average	Above Ave
MiG-19	15	15	4
MiG-21	3	15	16
MiG-23	9	11	14
F-4E	9	16	9
F-5E	6	18	10
Mirage 3C	3	16	15

Notice that the only difference between this ranking and the one obtained with paired comparisons is the reversal between the F-4E and F-5E. Using Spearman's rho again, this equates to a rank difference correlation of 0.983.

Using the derived category boundary values of -.839 and .467 (see Appendix III) it is interesting to note that all of the six systems scaled turn out to be "average". This information gives meaning to the scale values and suggests that the differences between the aircraft are slight. Figure 10 diagrams the final result.

To summarize, the successive interval method gives the analyst more information than either the paired comparison method or factor analysis without a reduction in precision. However, it cannot be used effectively to scale a group of weapons when there are <u>clearly</u> superior systems within the group because the dispersion of judgments will be too narrow to determine category bounds. In this sense it is not as robust as the previous methods. Nevertheless, in situations where the preferences among weapons are less obvious, it has great utility and helps the analyst interpret the scale values more readily.

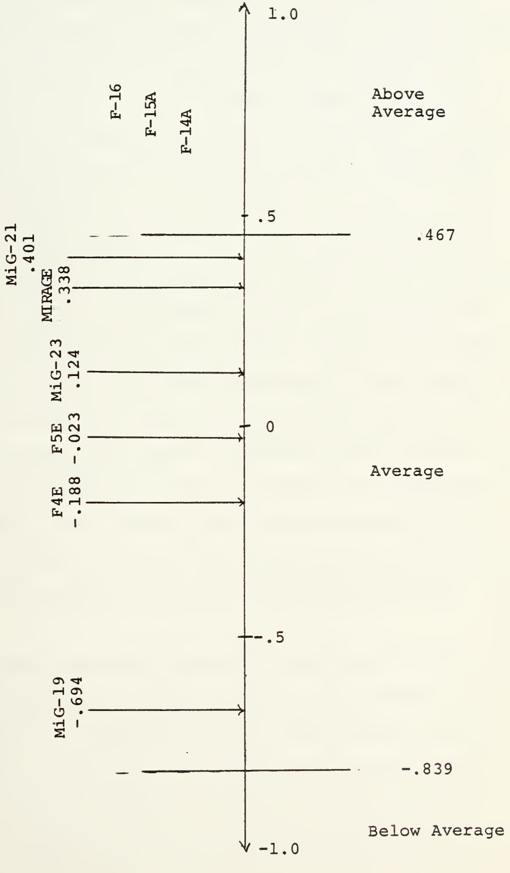
D. THE MULTI-ATTRIBUTE UTILITY APPROACH

1. Rationale

While the factor analytic, paired comparisons, and successive interval methods all provide interval-level capability measurement, it is often desirable to have even more

FIGURE 10

DIAGRAM OF SUCCESSIVE
INTERVAL RESULTS



precision. The most obvious need occurs when the analyst wishes to expand from a side-by-side comparison of weapons capability to a comparison of the military capability of one country with another. Recall, for example, the problems with Mihalka's country capability scores because of their derivation from interval data. Another useful calculation that cannot be performed with interval scores is percentage. This becomes important when the analyst needs to know _how much_ better one system is than another.

It is clear that _if_ things like country capability scores or percentage comparisons are desired, the researcher must strive for ratio measurement. None of the preceeding techniques can accommodate this demand. However, one possible avenue of approach that could produce a "conditional" ratio scale is multi-attribute utility theory (MAUT). It should be stressed that certain assumptions must be made to generate ratio information, some of which can be challenged. In the author's view, however, the assumptions are not unreasonable. Thus, as a potential approach to the ratio-data question, Multi-attribute utility theory warrants some attention.

A second important reason for investigating MAUT is that it accommodates an analysis of the multi-faceted definition of capability better than any of the previous techniques. Thus, even if its use as a ratio-measurement device can be challenged, it still merits consideration as an analytic technique.

2. Theory

Utility or value theory is a set of axioms designed to facilitate the decision process. Basically, it depends on the assumption that the decisionmaker will act rationally and always choose a course of action which maximizes expected utility (or usefulness) as defined by the decisionmaker's goals and the environmental constraints imposed upon him. When applied to the decision process, utility theory requires that all possible decision outcomes be quantified, the utility of each defined, and a decision reached based on maximizing utility.

Central to the entire process is the derivation of a utility function, u, which assigns a real value to each possible consequence such that the utility of consequence b, u(b), is greater than the utility of consequence c, u(c), if and only if consequence b is preferred to consequence c.[64] It is important to realize that the utility function depends on the subjective judgment of the decisionmaker and of his perception of the environment and the decision objectives. Equally important is the realization that, once defined, the utility function acts as an evaluative scale by which all possible outcomes are measured.

Until recently, most of the precepts in utility theory related to decisions which were based on a single

---

[64]Keeney, R.L., Multidimensional Utility Functions: Theory, Assessment and Application, p. 16, MIT, 1969.

attribute or criterion such as profit. As Keeney and others have pointed out, however, few decisions are based on just one measure of effectiveness.[65] This realization has prompted the development of multi-attributed utility theory and other procedures which attempt to cope with more complex problems. In view of the multidimensional nature of capability, this particular facet of general utility theory deserves further explanation.

According to Winterfeldt and Fischer, "multi-attribute utility theory (MAUT) combines a class of psychological measurement models and scaling procedures that can be applied to the evaluation of alternatives with multiple value relevant attributes."[66] For example, MAUT can be used to analyze preferences between cars described by the attributes cost, comfort, prestige, and performance. Similarly, it could be used to analyze weapons systems according to the series of specified attributes which define their capability.

A possible model for the employment of MAUT in weapon's capability assessment is suggested by the design engineering process advocated by design engineers to optimize system design and maximize system worth. As outlined by Kline and Lifson,[67] this process basically involves:

---

[65] Ibid., p. 9.

[66] Winterfeldt, D. and Fischer, G.W., Multi-Attribute Utility Theory: Models and Assessment Procedures, p. 1, NTIS, 1973.

[67] English, John M., (ed.), Cost-Effectiveness, The Economic Evaluation of Engineered Systems, pp. 79-80, John Wiley & Sons, 1968.

(1) Obtaining a clear statement of the goal or purpose of the system and the environment in which it is to operate

(2) selecting performance criteria which best define the objectives and assigning measures of utility to them to describe how valuable each criterion is

(3) comparing and weighting the various criteria to put their utility functions on a common basis or scale. (Once this is done, the individual utility functions can be combined into one objective function which can be used to calculate total system worth.)

(4) Using the output from steps 1 - 3, various alternatives are examined in light of estimated states of nature and an optimal solution is chosen based on trying to maximize expected utility.

While the purpose of the design engineer's analysis is finding an optimal system to fulfill an objective, the solutions obtained can also be used to evaluate or scale existing systems. In other words, the optimal solution can be considered as an ideal model against which all other systems can be judged.

This usage can be illustrated using fighter aircraft. Step 1 calls for a clear statement of the system's purpose and the operating environment. For illustrative purposes — and to be consistent with previous scaling done in the thesis — suppose the stated purpose is air-to-air combat. Additionally, consider the operational/technical environment to be similar to that found in the Middle East with the system designed to maximally perform at altitudes below twenty thousand feet AGL and at speeds between M 0.5 - 1.5. (The reasonableness of these parameters has been addressed previously.) These specifications direct and, in effect, <u>constrain</u> the judges to consider only what is important for maximizing air-to-air combat capability under these conditions. Thus, the analyst possessed a way to account for significant regional factors, tactics, or any other pertinent variables.

After clarifying the environment and purpose of the system, Step 2 calls for selecting criteria. As with Step 1, this can be done in general terms or in great detail; but whatever level is chosen, guidance should be obtained from experts (e.g., aeronautical engineers, pilots, etc.) or through a detailed technical analysis. The analysis performed in Chapter III provides the list of criteria used here.

Platform Criteria	Weapon Criteria	Miscellaneous Criteria
Max speed (energy)	Range	Technological level of the country
Acceleration (T/W)	Missile speed	Pilot proficiency
Maneuverability (W/S)	Firing envelope	
Endurance (combat radius)	No. of guns	

At this point the analyst must obtain from a series of subjective evaluations by his judge(s), a utility function defining the utility of each criterion over a range of values. For the capability problem and criteria tabulated above, the analysis must be done at two levels:

(1) Level I. Determine the individual utility function for each of the criteria under the three main dimensions, platform, weapon, and miscellaneous. For this particular problem this will lead to ten separate utility functions.

(2) Level II. Determine the utility relationships among the three dimensions as they relate to air-to-air combat capability.

The bi-level procedure thus leads to an over-all expression relating the utility of each criterion to the ultimate objective, air-to-air combat.

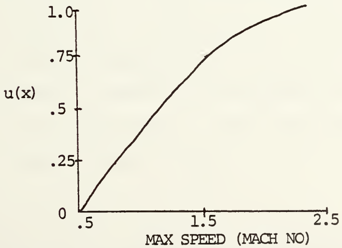
The actual process used to derive a utility function merits some attention. Consider, for illustrative purposes, maximum speed and define it as attribute x. The utility function, u(x), can be obtained over a range of values in the following way:

(1) First, determine the upper and lower limits (x^* and x_*) of the attribute, x. The lower limit, x_*, will be based on what the judge feels is the minimum speed required for an air-superiority aircraft. The upper limit,

x*, will be based on what the judge feels is technologically possible/desirable. By definition, the lower limit marks the threshhold where the utility, $u(x) = 0$, while the upper limit has a utility, $u(x) = 1.0$.

(2) Next, determine the general shape of the unidimensional utility function by defining the tradeoffs <u>within</u> each criterion. This is done by presenting the judge with a series of choices (lotteries) involving the criterion to be measured. (In the jargon of utility theorists, this is the same as finding out if there is "risk aversion," "increasing/decreasing risk aversion," etc. These topics are fully covered by Keeney.)[68]

(3) Finally, quantitatively assess the relative utilities of several speed values. This will result in a hypothetical curve such as:



---

[68] Keeney, R.L., <u>Multidimensional Utility Functions: Theory, Assessment and Application</u>, pp. 19-22, MIT, October, 1969.

Since the curve represents an expert's judgment of the utility of various speed capabilities, it can be used to scale the speed capability of existing systems. Thus, if an aircraft had a maximum speed of M1.5, it would have a utility of 0.75, based on the judge's preferences.

Once the utility functions for each of the criteria are derived, the important questions of how best to combine and weight each of the functions must be addressed. Put another way, the trade-offs among the criteria must be discerned. Weighting requires a judge's estimate of the relative importance of each criterion. The combination/simplification process is completely determined by the presence or absence of three properties within the utility functions, utility independence, pairwise preferential independence, and pariwise marginality.[69] These concepts can be defined as follows:[70]

(1) Utility independence. Assume consequence $\underline{x} = (x_1, \ldots, x_n)$ with utility $u(x)$. If $x_{\bar{i}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, x_i is utility independent of $x_{\bar{i}}$ if the decision-maker's relative preferences for x_i with $x_{\bar{i}}$ held fixed are the same regardless of the chosen value of $x_{\bar{i}}$.

---

[69] Giaque, William C., Prevention and Treatment of Streptococcal Sore Throat and Rheumatic Fever — A Decision Theoretic Approach, Ph.D. Thesis, p. IV-18, Harvard University, Nov. 1972.

[70] Ibid., pp. IV-15-19.

(2) Pairwise preferential independence is exhibited if the choice between two consequences $(x_1,x_2,x_3, \ldots, x_n)$ and $(x_a,x_b,x_3, \ldots., x_n)$ does not depend on the values of $x_3, \ldots, x_n$, for all pairs of attributes.

(3) Pairwise marginality holds if lottery (choice) $(x_i,x_j),(x_i^*,x_j^*)$ is indifferent to $(x_i,x_j^*),(x_i^*,x_j)$, where lottery A,B is a choice situation with the probabilities of consequences A and B both one-half.

Multiplication is called for if there is both utility independence and pairwise preferential independence. Under these conditions, the combined utility function, $u(\bar{x})$ becomes,[71]

$$1 + ku(\bar{x}) = \prod_{i=1}^{n} [1 + kk_i u_i(x_i)] \qquad (15)$$

where $u(\bar{x})$ is a multi-attributed utility function (e.g. platform), k and k_i are constants with $k > -1$ and $0 < k_i < 1$, $u_i(x_i)$ is the utility function of an individual criterion (e.g., speed), and π is the symbol for multiplication. Addition holds when utility independence, pairwise preferential independence, and pairwise marginality are all present. In this case the combined utility function, $u(\bar{x})$ is,[72]

---

[71] Ibid., p. IV-19.

[72] Ibid.

$$u(\overline{x}) = \sum_{i=1}^{n} k_i u_i(x_i) \qquad (16)$$

where $u(\overline{x})$ is again a multi-attributed utility function (e.g., platform, weapon, etc.), k_i is a constant, and $u_i(x_i)$ is the unidimensional utility function of a particular criterion (e.g., speed, thrust-to-weight ratio, etc.).

As with the individual criteria, the three structural attributes, platform, weapon, and miscellaneous, are also tested for utility independence, pairwise preferential independence, and pairwise marginality to determine their combinatorial relationship. The resulting equation will describe air-to-air combat in terms of utility values for each of the criteria which define it and will either be in the form of equation (15) or (16). At this stage, however, $u_i(x_i)$ is now a multi-dimensional function corresponding to either platform, weapon, or miscellaneous criteria elements, instead of just a unidimensional function for a single criterion.

It is possible that the requirements for addition or multiplication will not be present, or that they will not be decipherable. If such is the case most sources recommend uding the additive form since it is generally a good approximation of the multi-attribute utility function.[73]

---

[73]Keeney, op. cit., pp. 23-25.

One of the stated reasons for exploring MAUT was the chance that it could provide an acceptable ratio scale for capability. Assumptions are necessary because utility scaling is normally associated with interval measurement. In most applications this is an inescapable condition since the origin and unit are selected arbitrarily. Recall, however, that in the derivation of utility curves for capability assessment a zero-point is demanded from the judge and defined as the threshold value at which the attribute ceases to be useful to the specified objective. At least as far as the judge is concerned, therefore, this is an absolute zero-point. If the analyst is willing to generalize the validity of this point it can be considered a "natural" origin and lead to ratio measurement.

A second major assumption that must be made to allow ratio measurement is that the judge is capable of making reasonably precise judgments of the value of an attribute. Lifson argues that utility measurement is not exact.[74] People make mistakes and judgments are sometimes inconsistent. However, as demonstrated previously, judgmental measurement theory postulates a definite regularity in value judgments. On this basis, it can be argued that enough precision exists to measure at ratio level. (Some judgmental measuring techniques, e.g., the subjective estimate and constant sum methods,

---

[74] Lifson, Melvin W., Application of Criteria and Measures of Value in Engineering Design, Ph.D. Thesis, University of California, Los Angeles, p. 85, 1965.

actually produce ratio scales, for example.)[75] Nevertheless, the analyst must recognize the possible hazards involved in the assumptions as well as the fact that utility theory does not formally claim ratio measurement is possible.

3. Scaling Aerial Combat Capability with Multi-Attribute Utility Functions

To demonstrate the MAUT approach, utility curves were obtained for the ten criteria set forth previously. Two judges provided the necessary information. The first was an experienced fighter pilot with over 1000 hours in flight time and a graduate of the Navy Fighter Weapons School. His preferences are depicted in Figures 11-19. The second judge was a student at the Naval Postgraduate School with a degree in National Security Affairs. His judgments are represented in Figure 20, which scales the technological capacity of various countries.

Platform utility values were calculated for the nine aircraft rated in previous sections by evaluating their performance parameters against the utility functions in Figures 11-14. Weights (the k-values) were assinged to each of the criteria by the judge, and since none of the three simplification properties could be identified clearly, an additive solution was used. The values derived were:[76]

---

[75]Torgerson, op. cit., pp. 61-117.

[76]All calculations are included in Appendix IV.

FIGURE 11

UTILITY CURVE OF SPEED

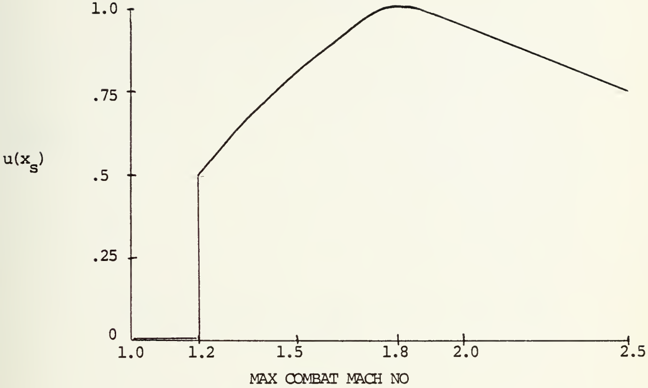


FIGURE 12

UTILITY CURVE OF T/W

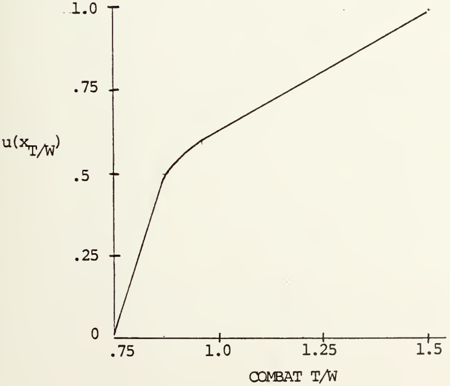


FIGURE 13

UTILITY CURVE FOR WING LOADING

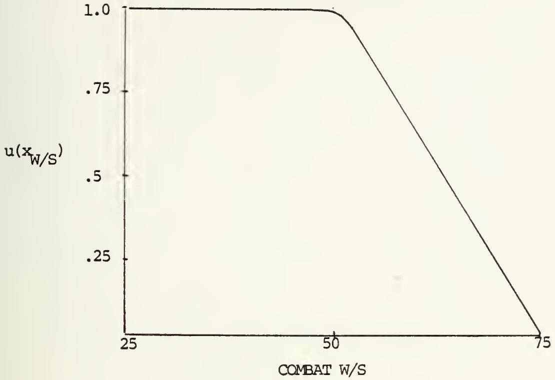


FIGURE 14

UTILITY CURVE FOR COMBAT RADIUS

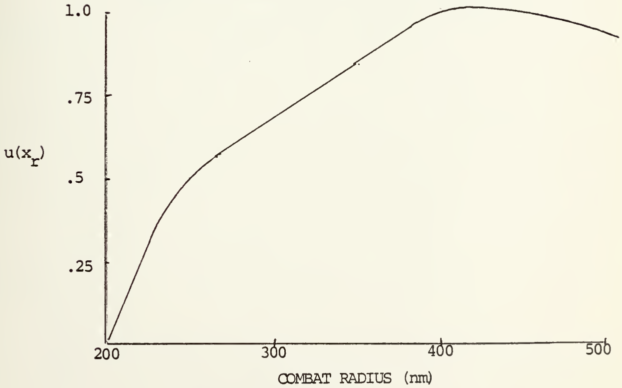


FIGURE 15

UTILITY CURVE FOR MISSILE RANGE

FIGURE 16

UTILITY CURVE FOR MISSILE SPEED

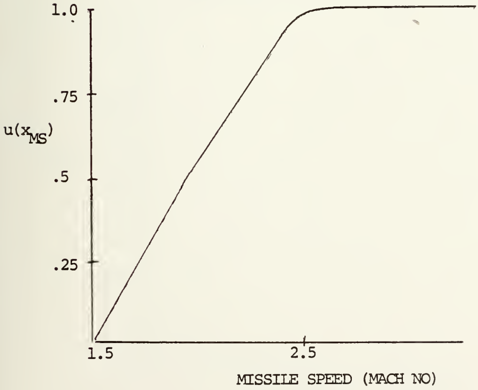


FIGURE 17

UTILITY CURVE FOR MISSILE FIRING ENVELOPE

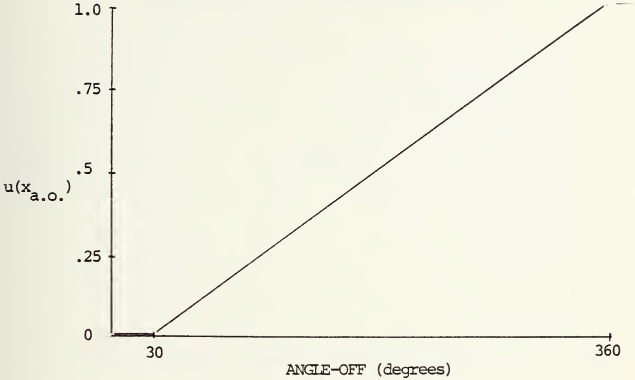


FIGURE 18

UTILITY CURVE FOR NO. OF GUN BARRELS

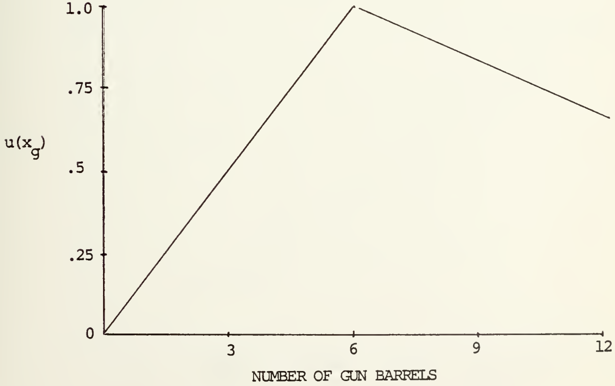


FIGURE 19

UTILITY CURVE FOR PILOT PROFICIENCY

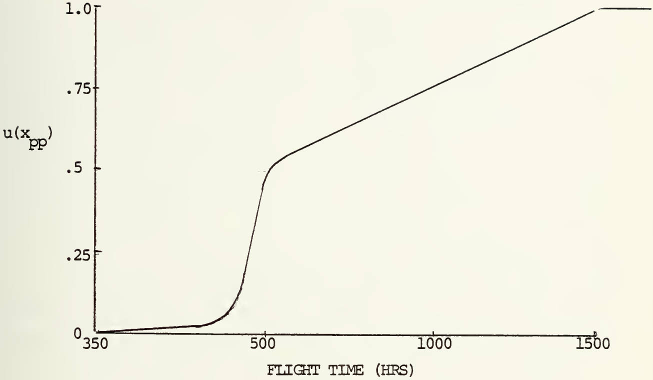
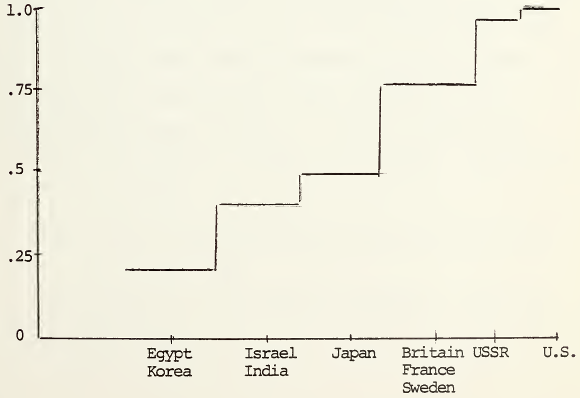


FIGURE 20

UTILITY CURVE FOR COUNTRY TECHNOLOGICAL CAPACITY



Aircraft	Platform Utility $u_p(x_p)$
F-16	.96
F-15	.86
F-14	.79
F-4E	.47
F-5E	.48
Mirage 3C	.69
MiG-19	.68
MiG-21	.62
MiG-23	.58

Once platform utility has been determined, the calculation is keyed to a particular country or situation by the nature of the remaining criteria. For example, the weapons utility score (derived from Figures 15-18) will depend on the country, tactics and situation under study. The same is true for the miscellaneous utility value since it embodies pilot proficiency and the technological capacity of a country.

To continue the calculation, some hypothetical data for Israel and Egypt were used to demonstrate the effects on system capability that accrue from a consideration of weapons and miscellaneous factors. The data assumed are:

(1) Two types of aircraft are present, the MiG-21 and F-4E

(2) Israeli pilots have an average of 1000 hours in flight time.

(3) Egyptian pilots have an average of 500 hours in flight time.

(4) All aircraft missile and gun systems are comparable.

Using the weights obtained from the judges (Appendix IV), and determining a multiplicative relationship to hold between platform utility, weapon utility, and miscellaneous utility, result in the F-4E having a utility value of .52 and the MiG-21 of .40. When compared to the original assessments of their platform utilities, these values are strikingly different. This suggests that while the MiG-21 may physically be a better aircraft than the F-4E (in terms of aerial combat), Israel's edge over Egypt in pilot proficiency and technical ability actually make the Israeli F-4E a superior system.

Obviously, such subtlety is not possible with any other technique discussed in this thesis. If reliable utility functions can be obtained, MAUT offers the greatest opportunity for realistic capability assessment since it takes into account more than just the weapons.

If the analyst can accept the assumptions needed for ratio measurement, these scores can be multiplied by inventory levels to obtain a measure of the usefulness of a country's fighter force in performing air-to-air combat. Simple arms transfer calculations can also be performed as well. For example, returning to the utility values just derived for the F-4E and MiG-21, if Israel had 50 F-4E's and Egypt 55 MiG-21's, the total air-to-air combat capability for each country — measured in <u>utils</u> — would be:

Israel: (.52) (50) = 26.0

Egypt: (.4) (55) = 22.0

Continuing, if the U.S.S.R. sent Egypt ten MiG-21's, the relative capabilities of the two countries (again measured in utils) would be the same. A supplier such as the U.S. would then have to send at least eight F-4E's to Israel to reinstate her previous capability advantage. It should be stressed that until the reliability of MAUT can actually be tested, it is not appropriate to advocate its wholesale adoption for measuring and comparing capabilities in this fashion. At this point, however, MAUT shows greater promise for ratio comparisons than any of the other techniques reviewed in this thesis.

V. CONCLUSIONS

The major concern driving this research has been the desire to improve the measurement and assessment of arms transfers. Several approaches are available, but one of the most valuable, and the one given most attention in this thesis, is capability assessment. Developing ways to measure capability has merit not only for political research, but for military intelligence estimates as well, and this broad usage gives impetus to the task.

In the course of this research, several important conclusions have been reached. First, the current use of factor analytic techniques to generate capability scores for weapons — and in particular, aircraft — should be reassessed. Not only has there been a tendency to oversimplify and misrepresent aircraft capability, but there is a tendency to misuse the scores once they have been derived. As interval-level data, factor scores can only be used for side-by-side comparisons of similar systems and cannot be used to calculate composite or country capability scores. Too much blind faith has been placed in the method because of its frequent us in data analysis problems and impressive theoretical framework, but not enough emphasis has been placed on the crucial tasks of selecting variables, interpreting factor results, and weighting and combining multiple-factor solutions. As this thesis has shown, without such guidance any one of a number of scores can be derived for a weapon which can change its comparative ranking with similar systems.

Reassessment does not mean abandonment. If decisions could be made, or at least a consensus reached, on some of these issues, factor analysis would become an extremely valuable tool for arms transfer measurement. Through SPSS it can bring the benefits of computer processing to bear on important research efforts.

Given the problems encountered in factor analysis, a second important conclusion reached in this thesis is that judgmental scaling techniques, in particular the paired comparisons and successive interval methods, are a viable alternative to factor analysis for capability assessment. In this research a high rank order correlation (rho = .983) was obtained when these techniques were used to scale aerial combat capability in nine modern fighter aircraft. More importantly, they eliminate the major problems encountered in factor analysis without sacrificing precision.

Of the two, it was discovered that the successive interval method was sensitive to certain distribution patterns in the judges' responses. If judgmental variation is less than three categories (intervals), boundaries and scores cannot be calculated. This says that interval scores cannot be determined for any system which is universally agreed upon as superior or inferior to another system. Although this restricts the application of the successive interval method to situations involving very similar systems, it is an extremely powerful method in that the analyst gains information on the meaning of the scores along with the scores themselves.

One glaring weakness this thesis uncovers in previous empirical studies which use capability scores is the tendency to use interval data as ratio-data. Many researchers would like to be able to measure capability on a ratio scale because it would allow absolute comparisons of the military potentials among countries. This thesis addresses, but does not adequately solve, measuring capability at a ratio-level. Multi-attribute utility theory provides the best avenue to a ratio scale, but certain assumptions must be made which reduce the authoritativeness of the results.

Ratio measurement aside, multi-attribute utility scaling is an impressive analytical approach which transforms human judgment into mathematical assessment. With respect to arms transfers, it is the only technique comprehensive enough to deal with capability as more than just a combination of performance characteristics. Hence, it should be given more than a cursory glance by military and political analysts.

If this work attests to anything, it is the fact that human judgment, and not a computer, is the key to capability assessment. The challenge to future analysis will be to make such judgments more precise by reaching agreement on what capability actually means. The two conceptual appraoches suggested in this thesis are a start in this direction.

APPENDIX I

Factor Analysis Results

Table A

Varimax Rotated Factor Matrix
For 29 Aircraft and 8 Variables

Variable	Factor I	Factor II	Factor III
Max Speed	.9486	.1412	.0962
Ceiling	.94610	-.1008	.0038
Combat Radius	-.25510	-.2630	.8137
Thrust-to-weight	.5923	.5762	-.0319
Wing loading	-.03127	-.8903	.0711
Number of gun barrels	-.0160	.8511	.2860
Missile Algorithm	.2292	.4231	.4999
Production Year	.2112	.3093	.7180

Table B

Oblique Rotated Factor Matrix
For 29 Aircraft and 8 Variables

Variable	Factor I	Factor II	Factor III
Max Speed	.0099	.9576	.0537
Ceiling	-.2343	.9835	-.0178
Combat Radius	-.2853	-.2133	.8502
Production Year	.2477	.1917	.6883
Thrust-to-weight	.5160	.5351	-.1021
Wing loading	-.9199	.0807	.1513
Gun Barrels	.8646	-.1175	.2130
Missile Algorithm	.3759	.1918	.4582

Factor Correlations

	Factor I	Factor II	Factor III
Factor I	1.0000	.2542	.1446
Factor II	.2542	1.0000	.0393
Factor III	.1446	.0393	1.0000

Table C

Rank Order for 12 Fighter Aircraft
Using Didderent Factor Combinations
(8 Variables – Oblique Rotation)

Factor I alone	Factor II+ alone
F-16	MiG-25
	F-15A
F-14A (with Phoenix	F-14A (with Phoenix)
F-15A	F-16
F-14A	MiG-21MF
MiG-23	MiG-21PF
Mirage 3C	F-14A
Mig-21 MF	MiG-23
MiG-21PF	Mirage 3C
F-4E	F-4E
MiG-19	MiG-19
F-5E	F-5E
MiG-25	

Factor I + Factor II	All Factors
F-15	F-14A (with Phoenix)
F-16	F-15A
F-14A (with Phoenix)	F-16
F-14A	F-14A
MiG-21MF	MiG-23
MiG-25	MiG-25
MiG-23	F-4E
Mirage 3C	Mirage 3C
MiG-21PF	MiG-21MF
F-4E	F-5E
F-5E	MiG-21PF
MiG-19	MiG-19

APPENDIX II

CALCULATIONS AND EQUATIONS FOR THE PAIRED COMPARISON METHOD

Equations for Scale Values
(Paired Comparisons)

(1.) $$S_{F16} - S_{F15} = \frac{1.732}{2} = .866$$

(2.) $$S_{F15} - S_{F14} = .593$$

(3.) $$S_{F14} - S_{M21} = \frac{3.084 + 1.658}{3} = 1.58$$

(4.) $$S_{M21} - S_{M23} = \frac{.43 + .509}{5} = .188$$

(5.) $$S_{M23} - S_{MIR-3C} = \frac{-.826 + .704}{5} = -.0244$$

(6.) $$S_{MIR-3C} - S_{F4E} = \frac{.948 - .173}{4} = .194$$

(7.) $$S_{F4E} - S_{F5E} = \frac{-.155 + 1.142}{4} = .247$$

(8.) $$S_{F5E} - S_{M19} = \frac{-1.77 + 1.855}{4} = .021$$

FIGHTER CAPABILITY MEASUREMENT

The purpose of this questionnaire is to gather data on fighter aircraft capability for a project on arms transfers currently under study by JCS and members of the Government Department at NPS. For each of the 36 pairs of aircraft listed below circle the one you consider to be superior in terms of aerial combat capability. Consider each pair separately. Assume a combat environment between 5-20K ft and speeds between M 0.5-1.5. Also assume that no long-range missiles such as Phoenix are available and that pilots are approximately of equal ability.

Some performance data has been provided to insure a certain level of knowledge for unfamiliar systems. However, your judgment/experience is valued more since the data was taken from open sources and may not be exact.

1.	F-4E	F-16	19.	Mirage 3C	MiG-23
2.	MiG-21	F-5E	20.	F-5E	MiG-23
3.	MiG-19	F-14A	21.	MiG-19	F-5E
4.	F-15A	F-4E	22.	MiG-19	MiG-23
5.	MiG-21	F-4E	23.	F-15A	F-16
6.	MiG-23	F-15A	24.	F-16	MiG-23
7.	F-5E	F-16	25.	F-5E	F-14A
8.	Mirage 3C	F-5E	26.	F-4E	MiG-19
9.	F-15A	MiG-19	27.	F-14A	MiG-23
10.	F-15A	F-5E	28.	F-15A	MiG-21
11.	F-16	Mirage 3C	29.	F-4E	Mirage 3C
12.	F-14A	Mirage 3C	30.	F-4E	F-14A
13.	MiG-19	F-16	31.	F-4E	F-5E
14.	MiG-19	Mirage 3C	32.	F-15A	F-14A
15.	F-16	MiG-21	33.	Mirage 3C	MiG-21
16.	MiG-21	MiG-23	34.	MiG-19	MiG-21
17.	F-15A	Mirage 3C	35.	F-4E	MiG-23
18.	F-14A	F-16	36.	F-14A	MiG-21

APPENDIX III

CALCULATIONS AND EQUATIONS FOR THE SUCCESSIVE INTERVAL METHOD

I. Solution to the law of categorical judgment (equation 13)

1. Given the equation

$$z_{ig} = \frac{B_g - S_i}{\sigma_i^2 + c} \qquad (1)$$

2. If the estimates of z_{ig} are added over instances (column sums), equation 1 becomes

$$\sum_{i=1}^{n} z_{ig} = B_g\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + c}\right) - \left(\sum_{i=1}^{n} \frac{S_i}{\sigma_i^2 + c}\right) \qquad (2)$$

3. Think of equation (2) as a linear transformation of an interval scaled variable. The first term on the right-hand side, $\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + c}$, establishes the unit of the scale, while the second term, $\sum_{i=1}^{n} \frac{S_i}{\sigma_i^2 + c}$, establishes the origin. Since the unit and origin can be arbitrarily set, let the origin, $\sum_{i=1}^{n} \frac{S_i}{\sigma_i^2 + c} = 0$, and let the unit,

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + c} = n.$$

4. Using these relationships reduces equation 2 to

$$\sum_{i=1}^{n} z_{ig} = n\, B_g, \quad \text{or} \quad \sum_{i=1}^{n} \frac{z_{ig}}{n} = B_g \qquad (3)$$

5. Equation 3 says that an estimate of the category upper bound, B_g, is obtained from the column average of the z-score array.

6. Since estimates for z_{ig} and B_g now exist, the only thing required to solve equation 1 is an estimate of the variance, $\frac{1}{\sigma_i^2 + c}$. To do this, it is first necessary to find the row average, $\bar{z}_i$. If there are m+1 categories,

$$\bar{z}_i = \sum_{g=1}^{m} \frac{z_{ig}}{m} = \frac{1}{\sigma_i^2 + c} \left(\sum_{g=1}^{m} \frac{B_g}{m} - \sum_{g=1}^{m} \frac{S_i}{m} \right)$$

$$\bar{z}_i = \frac{1}{\sigma_i^2 + c} [\bar{B}_g - S_i] \qquad i = 1, 2, \ldots, n. \qquad (4)$$

where B_g is the average of the column averages.

7. From equations 1 and 4

$$z_{ig} - z_i = \frac{B_g - \bar{B}}{\sigma_i^2 + c}$$

8. Squaring both sides, adding over categories, and rearranging terms

$$\sigma_i^{2} + c = \frac{\sum_{g=1}^{m} (B_g - \bar{B})^2}{(z_{ig} - \bar{z}_i)^2} \qquad \begin{array}{l} i = 1, \ldots, n \\ g = 1, \ldots, m \end{array} \quad (5)$$

9. Since $S_i = B_g - z_{ig} \quad \sigma_i^{2} + c$

$$S_i = B_g - z_{ig} \frac{\sum_{g=1}^{m} (B_g - \bar{B})^2}{\sum_{g=1}^{m} (z_{ig} - \bar{z}_i)^2} \qquad \begin{array}{l} i = 1, \ldots, n \\ g = 1, \ldots, m \end{array}$$

10. Or, if A is the sum of the squares of the columns, and C_i is the sum of the squares of the rows, the scale value of each instance (row) is,

$$S_i = B - \bar{z}_i \sqrt{\frac{A}{C_i}}$$

II. The solutions identified in Part I are now used to obtain the results reported in Chapter IV.

1. Step 1: Arrange raw frequency data in a table where rows are instances and columns are categories, with column 1 being least favorable, etc.

MiG-19	15	15	4
MiG-21	3	15	16
MiG-23	9	11	14
F-4E	9	16	9
F-5E	6	18	10
Mirage-3C	3	16	15

2. Step 2: Compute the relative cumulative frequencies for each row and record them in a new table. Exclude the last column since it will be a unit column vector.

MiG-19	.441	.882
MiG-21	.088	.529
MiG-23	.265	.588
F-4E	.265	.735
F-5E	.176	.706
Mirage-3C	.088	.559

3. Step 3: Treating these values as leftward areas under a N(0,1) curve, find the standard normal deviates for these areas and record them in another table. This array will have one less column than the original array in step 1.

MiG-19	-.148	1.186
MiG-21	-1.354	.073
MiG-23	-.625	.223
F-4E	-.625	.628
F-5E	-.931	.542
Mirage-3C	-1.354	.148

4. Step 4: Compute a row average for each row, a column average for each column, a grand average, $\overline{B}$, and the sum of the squares of the columns, A.

				Row Sum	$\overline{z}_i$
MiG-19	- .148	1.186	=	1.038	.519
MiG-21	-1.354	.073	=	-1.281	-.641
MiG-23	- .625	.223	=	- .402	-.201
F-4E	- .625	.628	=	.003	.002
F-5E	- .931	.542	=	- .389	-.184
Mirage-3C	-1.354	.148	=	1.206	-.603

Column Sum = -5.037 + 2.8 = > Grand Average $\overline{B} = -.186$

B_g = - .839 .467

$\uparrow$ B_1 $\uparrow$ B_2

$$A = (-.839 + .186)^2 + (.467 + .186)^2 = .852$$

5. Step 5: For each row compute C_i.

$$C_1 = \text{Mig-19} = (-.148 - .519)^2 + (1.186 - .519)^2 = .890$$

$$C_2 = \text{MiG-21} = (-1.354 + .641)^2 + (.073 + .641)^2 = 1.018$$

$$C_3 = \text{MiG-23} = (-.625 + .201)^2 + (.223 + .201)^2 = .360$$

$$C_4 = \text{F-4E} = (-.625 - .002)^2 + (.628 - .002)^2 = .785$$

$$C_5 = \text{F-5E} = (-.931 + .184)^2 + (.542 + .184)^2 = 1.085$$

$$C_6 = \text{Mirage-3C} = (-1.354 + .603)^2 + (.148 + .603)^2 = 1.128$$

6. Step 6: For each row compute $\sqrt{\frac{A}{C_i}}$. This gives an estimate of the variance.

$$\text{MiG-19} = \frac{.852}{.890} = .978$$

$$\text{MiG-21} = \frac{.852}{1.018} = .915$$

$$\text{MiG-23} = \frac{.852}{.360} = 1.54$$

$$\text{F-4E} = \frac{.852}{.785} = 1.042$$

$$\text{F-5E} = \frac{.852}{1.085} = .886$$

$$\text{Mirage-3C} = \frac{.852}{1.128} = .869$$

7. Step 7: Compute the scale values, S_i, according to the equation, $S_i = \bar{B} - \bar{z}_i \sqrt{\frac{A}{C_i}}$

MiG-19 = -.186 - .519(.978) = -.694

MiG-21 = -.186 + .641(.915) = .401

MiG-23 = -.186 + .201(1.54) = .124

F-4E = -.186 - .002(1.042) = -.188

F-5E = -.186 + .184(.886) = -.023

Mirage-3C = -.186 + .603(.869) = .338

8. Step 8: The scale can be transformed by a linear transformation to best suit the needs of the analyst. In this instance all values are transformed so that the MiG-19 has a score of 1.0 by adding 1.694 to every score.

MiG-21 = 2.095

Mirage 3-C = 2.032

MiG-23 = 1.818

F-5E = 1.671

F-4E = 1.506

MiG-19 = 1.000

FIGHTER CAPABILITY MEASUREMENT

The purpose of this questionnaire is to gather data on the capability of fighter aircraft. You are asked to categorize each of the nine aircraft listed below into one of five categories – excellent, above average, average, below average, poor – on the basis of their serial combat capability. Assume a combat environment of 5-20K ft with speeds ranging between M.5 - 1.5. Also assume that no long range missiles are available and that aircraft are piloted by men of approximately the same ability.

If in doubt about the performance of a certain aircraft, use the data provided. However, your judgment/experience is valued more since the data was taken from open sources and may not be exact.

	EXCELLENT	ABOVE AVERAGE	AVERAGE	BELOW AVERAGE	POOR
MiG-19					
MiG-21 PFM					
MiG-23					
F-4E					
F-5E					
F-14					
F-15					
F-16					
Mirage-3C					

APPENDIX IV

Multi-Attribute Utility Calculations

I. Level I Calculations

A. Determination of platform utility $u_p(\bar{x}_p)$.

Weights assigned by judge for platform criteria:

$$k_s = .15 \qquad k_{w/s} = .35$$

$$k_{t/w} = .40 \qquad k_r = .10$$

Additive function is assumed, $u_p(\bar{x}_p) = \sum_{i=1}^{n} k_i \, u_i(x_i)$, and $\sum_{i=1}^{n} k_i = 1$. Therefore, entering the utility curves with the weapons performance characteristics for each aircraft:

values from utility curves

F-16 = $(1.0)(.15) + (.9)(.4) + (.35)(1.0) + (.1)(1.0) = \underline{0.96}$

F-15A = $(.88)(.15) + (.8)(.4) + (.35)(.9) + (.95)(.1) = \underline{.86}$

F-14A = $(.95)(.15) + (.5)(.4) + (.35)(1.0) + (.1)(.95) = \underline{.79}$

F-4E = $(.84)(.15) + (.6)(.4) + (.35)(0) + (.1)(1.0) = \underline{.47}$

F-5E = $(.63)(.15) + (.25)(.4) + (.35)(.6) + (.1)(.8) = \underline{.48}$

Mir.3C = $(.95)(.15) + (.25)(.4) + (.35)(1.0) + (.1)(1.0) = \underline{.69}$

MiG-19 = $(.5)(.15) + (.55)(.4) + (.35)(.88) + (.1)(.8) = \underline{.68}$

MiG-21 = $(.62)(.15) + (.5)(.4) + (.35)(.8) + (.1)(.5) = \underline{.62}$

MiG-23 + $(.75)(.15) + (.5)(.4) + (.35)(.5) + (.1)(.9) = \underline{.58}$

B. Determination of weapon utility, $u_w(\overline{x}_w)$.

Weights assigned by judge for weapons criteria:

$k_{mr} = .15 \qquad k_{ms} = .15$

$k_{a.o.} = .30 \qquad k_g = .40$

For Israeli/Egyptian calculation assume

$$u_w(\overline{x}_w) = k_i\ u_i(x_i) = \underline{.5}$$

C. Miscellaneous Utility, $u_m(\overline{x}_m)$.

Weights assigned by judge:

$k_{country} = .2$

$k_{pilot} = .8$

Additive function holds, therefore,

$$u_m(x_m) = k_{country}\ u(x)_{country} + k_{pilot}\ u(x)_{pilot}$$

Israel $u_m = .2(.35) + .8(.75) = .67$

Egypt $u_m = .2(.2) + .8(.5) = .44$

II. Level II Calculation for over-all aerial combat utility

A. Multiplicative relationship was determined. Therefore:

$$1 + k\ U_A(X_A) = [1 + kk_p u_p(x_p)] \times [1 + kk_w u_w(x_w)] \times [1 + kk_m u_m(x_m)]$$

over-all aerial combat utility	=	platform utility	x	weapons utility	x	miscellaneous utility

B. Weights assigned by judge:

$$k_p = .2 \qquad k_w = .1 \qquad k_m = .6$$

C. From Level I analysis:

$u_p(x_p)$ for F-4E = .47 $\qquad$ $u_p(x_p)$ for MiG-21 = .62

$u_w(x_w)$ = .5 (assumed) $\qquad$ $u_m(x_m)$ for Israel = .67

$u_m(x_m)$ for Egypt = .44

D. Solving for k: Let $U_A(X_A) = 1$; therefore,

$$u_p(x_p) = u_w(x_w) = u_m(x_m) = 1.$$

$$1 + k = (1 + .2k)(1 + .1k)(1 + .6k)$$

$$.012\ k^3 + .2\ k^2 - .1\ k = 0$$

$$k = -.42 \text{ and } -16.25$$

always use the root between -1 and ∞

$$k = \underline{-.42}$$

E. Substituting the value of k in the utility expression (II-A) along with the values from II-B and II-C:

Israeli F-4E:

$$1 - .42[U_A(X_A)] = [1 - .084(.47)][1 - .042(.5)][1 - .25(.67)]$$

$$U_A(X_A) = \underline{.52}$$

Egyptian MiG-21:

$$1 - .42[U_A(X_A)] = [1 - .084(.62)][1 - .042(.5)][1 - .25(.44)]$$

$$U_A(X_A) = \underline{.40}$$

BIBLIOGRAPHY

Advisory Group For Aeronautic Research and Development (AGARD) Conference Proceedings (Various Issues)

Anderberg, Michael R., Cluster Analysis for Applications, Academic Press, 1973.

Arms Control and Disarmament Agency, The International Transfer of Conventional Arms, ACDA, 1973.

Aviation Week and Space Technology (Various Issues)

Babister, A.W., Aircraft Stability and Control, Parmajon Press, 1961.

Blalock, Hubert M., Social Statistics, 2nd ed., McGraw-Hill, 1972.

Bloomfield, Lincoln P. and Leiss, Amelia C., Controlling Small Wars, Alfred A. Knopf, 1969.

Casey, Glen A., A Methodological Approach for Analyzing South American Arms Acquisition and U.S. Policy, Master's Thesis, Naval Postgraduate School, Monterey, California, 1974.

Churchman, C.W. and Ratoosh Philburn (ed.), Measurement Definitions and Theories, John Wiley and Sons, 1959.

Coleman, James S., Policy Research in the Social Science, General Learning Press, 1972.

Dennis, Michael F.H., "The Policy Basis of General Program Forces and A Model for Quantitative Analysis," The Journal of Conflict Resolution, vol. xviii, Number 1, March 1974.

Dommasch, D.O., Sherby, S.S., and Connelly, Thomas, T., Aeroplane Dynamics, 3rd Ed., Pitman Publishing, 1961.

Duran, Benjamin S. and Odell, Patrick L., Cluster Analysis: A Survey, Springer-Verlag, 1974.

English, J. Morley (ed.), Cost Effectiveness: The Economic Evaluation of Engineered Systems, John Wiley and Sons, 1968.

Everitt, Brian, Cluster Analysis, John Wiley and Sons, 1974.

Frank, Lewis A., The Arms Trade in International Relations, Praeger, 1969.

Garson, G. David, Handbook of Political Science Methods, Holbrook Press, 1971.

Giaugue, William C., Prevention and Treatment of Streptococcal Sore Throat and Rheumatic Fever — A Decision Theoretic Approach, Ph.D. Thesis, Harvard University, 1972.

Gilbert, Stephen P. and Joshua, Wynfred, Arms for the Third World, Johns Hopkins Press, 1969.

Green, Paul E. and Carmone, Frank J., Multidimensional Scaling and Related Techniques in Marketing Analysis, Allyn and Bacon, Inc., 1970.

Green, Ted Robert, Politmetrics, An Introduction to Quantitative Macropolitics, Prentice-Hall, 1972.

Harkavy, Robert E., Arms Transfers and the International System, paper presented at the 32nd annual meeting of the Midwest Political Science Association, April, 1974.

Hoagland, John H., "Arms in the Developing World," Orbis, September , 1968.

Hurewitz, J.C. (ed.), Soviet American Rivalry in the Middle East, Praeger, 1969.

Insight Team of the London Sunday Times, The Yom Kippur War, Doubleday, 1974.

Jane's All The Worlds Aircraft, (Various Issues)

Keeney, R.L., Multidimensional Utility Functions: Theory Assessment and Applications, Technical Report No. 43, MIT, October, 1969.

Kemp, Geoffrey and Sutton, J.R., Arms to Developing Countries 1945-1965, Adelphi Paper No. 28, International Institute for Strategic Studies, 1966.

Kemp, Geoffrey, Classification of Weapons Systems and Force Designs in Less Developed Country Environments: Implications for Arms Transfer Policies, C/70-3, MIT, 1970.

Kemp, Geoffrey, "Arms Traffic to Third World Conflicts", International Conciliation, March, 1970.

Kemp, Geoffrey, "Strategy, Arms, and the Third World: A Review of the Arms Trade With The Third World (SIPRI)", Orbis, Fall, 1972.

Lambelet, John C., "A Dynamic Model of the Arms Race in the Middle East, 1953-1965", General Systems, Vol. XVI, 1971.

Learmonth, G.P., Technical Note No. 0211-19, Directions for Using SPSS, Naval Postgraduate School, Monterey, California 1974 (mimeo).

Leiss, Amelia C. et. al., Arms Transfers to Less Developed Countries, C/70-1, MIT, 1970.

Leiss, Amelia C., Changing Patterns of Arms Transfers, Implications for Arms Transfer Policies, C/70-2, MIT, 1970.

Lifson, Melvin W., Application of Criteria Measures of Value in Engineering Design, Ph.D. Thesis, University of California, Los Angeles, 1965.

Maranell, G.M. (ed.), Scaling: A Sourcebook for Behavioral Scientists, Aldine Publishing, 1974.

McDonnell Douglas, Tactical Manual Navy Model F-4B and F-4J Aircraft (u), NAVAIR 01-245 FDB-1T 1 March 1971.

Mihalka, Michael, Understanding Arms Accumulation: The Middle East as an Example, The University of Michigan (mimeo), March, 1973.

Naylor, J.L., Dictionary of Aeronautical Engineering, Philosophical Library, 1959.

Nie, Norman et. al., Statistical Package for the Social Sciences, 2nd ed., McGraw-Hill, 1975.

Rosen, S.J. and Kurth, J.R., Testing Theories of Economic Imperialism, D.C. Heath, 1974.

Rummel, Rudolph J., "Understanding Factor Analysis", Journal of Conflict Resolution.

Rummel, Rudolph J., Applied Factor Analysis, Northwestern University Press, 1970.

Shepard, R.N., Romney, A.K. and Nerlove, S.B., (ed.) Multidimensional Scaling, Vol. I, Seminar Press, 1972.

SIPRI Yearbook of World Armaments and Disarmaments, Humanities Press, 1968/69, 1969/70, 1972, 1973, 1974, 1975.

(SIPRI) Stockholm International Peace Research Institute, The Arms Trade With The Third World, Humanities Press, 1971.

Sivard, Ruth L., World Military and Social Expenditure, WMSE Publications, 1974.

Snider, Lewis W., and Perkins, L., Armtran II: Arms Transfers to Less Developed Countries, A Codebook, University of Michigan (mimeo), February 1974.

Snider, Lewis W., Arabesque: Untangling the Patterns of Supply of Conventional Arms to Israel and the Arab States and Their Effects on the Arab/Israeli Conflict 1948-1973, paper presented at the 1975 Annual Convention of the International Studies Association, Washington, D.C., 19-22 February 1975.

Snider, Lewis W., Middle East Laestrom. The Impact of Global and Regional Influences in the Arab-Israeli Conflict 1947-1973, Ph.D. Thesis, University of Michigan, 1975.

Space/Aeronautics, (Various Issues)

Stanley, John and Pearton, Maurice, The International Trade in Arms, Praeger, 1972.

Stevens, S.S., Handbook of Experimental Psychology, Wiley, 1951.

Thayer, George, The War Business: The International Trade and Armaments, Simon and Schuster, 1969.

Toyerson, Warren S., Theory and Methods of Scaling, John Wiley and Sons, 1967.

Tufte, Edward R. (ed.), The Quantitative Analysis of Social Problems, Addison Wesley, 1970.

Tufte, Edward R., Data Analysis for Politics and Policy, Prentice-Hall, 1974.

Wilson, Richard, V.N., Escort Cost Model, Center For Naval Analyses, October 1973.

Winterfeldt, D. and Fischer G.W., Multi-Attribute Utility Theory: Models and Assessment Procedures, Technical Report, University of Michigan, 5 November 1973.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 56 Department of National Security Affairs Naval Postgraduate School Monterey, California 93940	2
4.	Assoc. Professor E.J. Laurance, Code 56Lk Department of National Security Affairs Naval Postgraduate School Monterey, California 93940	1
5.	Asst. Professor R.G. Sherwin, Code 56 Department of National Security Affairs Naval Postgraduate School Monterey, California 93940	1
6.	Commander, Naval Intelligence Command 2461 Eisenhower Avenue Alexandria, Virginia 22331	2
7.	LT Allan W. LeGrow Commander Operational Test and Evaluation Force Norfolk, Virginia 23511 ATTN: Code 24	1
8.	Major Michael J. Hanley, USMC Strategic Plans and Policy Division (B215) The Pentagon Washington, D.C. 20301	1